

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problems Mailbox.**

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



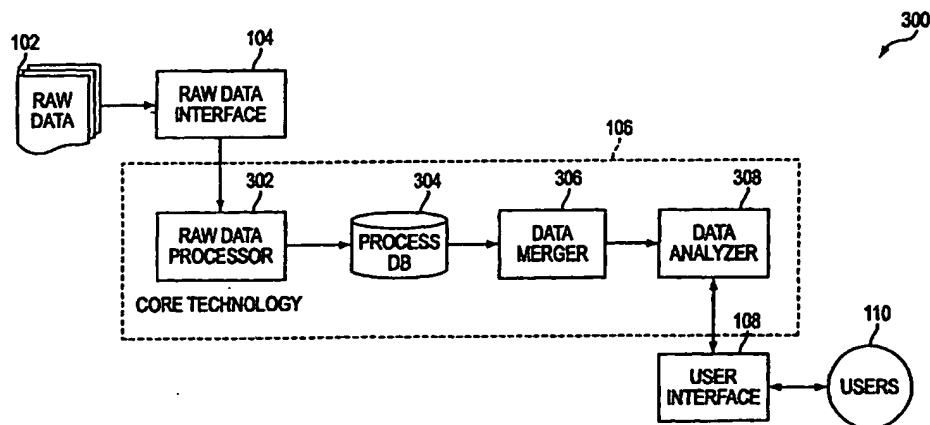
(43) International Publication Date
28 December 2000 (28.12.2000)

PCT

(10) International Publication Number
WO 00/79449 A2

- (51) International Patent Classification⁷: G06F 17/60
- (21) International Application Number: PCT/US00/15823
- (22) International Filing Date: 9 June 2000 (09.06.2000)
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data:
09/328,898 9 June 1999 (09.06.1999) US
09/379,587 24 August 1999 (24.08.1999) US
- (71) Applicant: FUTURISTICS LABS, INC. [US/US]; 9500 Brooktree Road, Suite 200, Wexford, PA 15090 (US).
- (72) Inventors: KIRBY, Christopher, M.; 4621 Sutton Place Ext, Wexford, PA 15090 (US). CHANG, Steven, C., P.; 5224 Karrington Drive, Gibsonia, PA 15044 (US). BARTELS, John, D.; 825 South Negley Avenue, #3, Pittsburgh, PA 15232 (US).
- (74) Agent: VENABLE; P.O. Box 34385, Washington, DC 20043-9998 (US).
- (81) Designated States (*national*): AE, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZA, ZW.
- (84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).
- Published:
— Without international search report and to be republished upon receipt of that report.
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: SYSTEM, METHOD AND COMPUTER PROGRAM PRODUCT FOR GENERATING AN INVENTORY-CENTRIC DEMOGRAPHIC HYPER-CUBE



(57) Abstract: A method, system, and computer program product for analyzing client user accesses to the Internet in a substantially real-time manner can include accessing raw data, processing the raw data using a core technology, and interfacing the raw data and core technology using a virtual cookie to obtain clean raw data. Interfacing includes accessing a proxy log and an IP address assignment log, and merging the logs to obtain virtual cookie identification clean raw data. Processing using core technology includes receiving clean raw data, processing it using a raw data processor, processing the output to obtain an inventory-centric demographic hyper-cube (cube), merging a plurality of cubes into a merged cube, and analyzing the merged cube. Processing the raw data processor output includes loading user demographics and actions, detecting and removing robots, determining behavioral interest groups and user profiles, and building the cubes.

WO 00/79449 A2

System, Method and Computer Program Product for Generating an Inventory-Centric Demographic Hyper-Cube

Background of the Invention

Field of the Invention

The present invention relates to understanding purchasing behavior in traditional stores and to an improved method of performing demographic, psychographic, and behavioral analysis of the Internet.

Cross-Reference to Related Applications

U.S. Patent Application Serial No. 09/277,751, entitled "System, Method and Computer Program Product for Creating a Virtual Cookie," filed March 29, 1999, by Messrs. C. M. Kirby and S. C. P. Chang, of common assignee, U.S. Patent Application Serial No. 09/328,898, entitled "System, Method and Computer Program Product for Generating an Inventory-Centric Demographic Hyper-Cube," filed June 9, 1999 to Messrs. C. M. Kirby, S. C. P. Chang and J. D. Bartels, of common assignee, and U.S. Patent Application Serial No. 09/379,587, entitled "System and Method and Computer Program Product for Reporting User Behavior Statistics," filed August 24, 1999, by Messrs. C. M. Kirby and S. C. P. Chang, of common assignee to the present invention, the contents of which are incorporated herein by reference in their entirety.

Related Art

Whenever there is a lot of user activity, such as for example, visitors interacting with a web site or customers making purchases in a store, there is a strong desire to understand the behavior of the users. User behavior can be used to better target advertising or to find potential buyers for a product or service. Transforming large volumes of raw data regarding user activity into an understanding of expected user behavior is a continuing technical challenge.

Efforts have been attempted to meet this challenge. For example, a first technique for solving this problem can involve simple counting of attributes of interest. Specific examples can include tracking the amount of user activity, the number of users, the number of users that

visited a given page X, and the breakdown of males versus females. Companies like WebTrends of Portland, OR and others perform this type of analysis. By providing these metrics, this technique can increase understanding by providing information on the total audience. Unfortunately, this technique does not scale well when information on targeted audiences rather than a total audience is needed.

A second technique involves user-centric clustering that extends the first technique to provide information on subgroups of the total audience. All users can be assigned to a cluster, using a classification, such as, e.g., matching pre-defined categories, or using traditional clustering techniques, where clusters are generated dynamically. Personify of San Francisco, CA and DataSage of Reading, MA are examples of companies performing this type of analysis. With users assigned to clusters, the second technique can generate totals as before, but on a per-cluster basis. So, e.g., if there are five clusters of users, then there could be five sets of totals. This technique can allow for an understanding of subgroups of the total audience. Unfortunately, the second technique is constrained in that it can only provide information about those users represented by clusters. If there is interest in some other subgroup not represented by a cluster, this second technique cannot offer any information.

Increased use of the global Internet has created a need for improved identification, tracking and analysis of web server access by client users. Advertisers, e.g., are interested in targeting ads to particular users. Electronic commerce (e-commerce) companies also attempt to target customers on the Internet. Web servers also want to recognize a return visitor to a web site in order to provide customized presentation of a web page. Different methods of detecting and tracking access to web sites are available. However, conventional web traffic analysis and tracking tools have limitations.

Effectiveness of conventional tracking and reporting systems analyzing user accesses of web sites is limited by the granularity of data available regarding user behavior, and by methods used to access the Internet.

Conventional user interfaces for analyzing demographic data are limited in that they provide statistical summary data only on a per site basis. It is desirable that statistics regarding user behavior such as, e.g., Internet user behavior, be provided on a per location basis. Unfortunately, conventional systems cannot provide user behavior information on a per location

basis. Instead, conventional systems can only provide user traffic statistics on a per-site basis. A per site basis, as compared to a per location basis, can only provide statistics regarding Internet user behavior traffic generally about a site itself, and cannot provide granular statistical data down to the level of a specific page within that site.

The term "location," as used in the present invention, refers to a distinct webpage within a website. For example, a website of Amazon.com can include many web pages, each web page corresponds to a separate file having a separate universal resource locator (URL) filename associated with it. Each individual web page URL filename of the Amazon.com site, such as, e.g., <http://www.amazon.com/subdirectory/subsubdirectory/filename.htm>, is thus referred to as a "location." User behavior statistics are not conventionally available to this level of granularity.

Analysis of behavior of client users can include tracking demographic attributes of the users of the Internet. A demographic attribute can include a "pure demographic" attribute such as, e.g., a client user's age, gender, and salary range. Another demographic attribute can be obtained by analyzing the behavior of a client user. It is desirable that user demographic statistical data be available on a per location basis.

Many Internet clients access the Internet by using proxy servers and network communications servers (NCSs). Internet service providers (ISP) often use proxy servers and NCSs.

Proxy servers can shield some Internet requests by a client host from the rest of the Internet. For example, proxy servers often cache, i.e., store for future access, certain popular web pages from a web server. Caching can improve access time to the web page for users and can save communications costs for the ISP. When the web client requests access to a cached web page, the cached web page is accessed from the proxy server's cache and no request is made to the web site where the requested web page resides. It is possible then that the web server of the requested page is never accessed once the page is cached. Thus the web server is not made aware that the web client accessed its web server site.

All Internet hosts, including both clients and servers, must have their own Internet protocol (IP) addresses. An Internet host's IP address is analogous to a postal mailing address and is used for sending information between multiple Internet hosts. A network communications

server (NCS) is often used to assign an IP address to a computer host. The NCS can permanently assign an IP address to a host, known as static IP address assignment. The NCS can also temporarily assign IP addresses, known as dynamic IP address assignment.

Web servers have, for a long time, had the ability to customize a web site for a particular person on a person-by-person basis. Imagine how difficult it would be to maintain a list of preferences for each user that ever visited a particular search site such as, e.g., Yahoo. To keep such preferences up-to-date, if a Yahoo web server was being accessed by millions of users, then it could amount to millions of bytes of data requiring to be stored on the web server, which would need to be retrieved in a timely manner. It was thought to be better to have each user maintain his or her own preferences locally to eliminate retrieval time and to maintain privacy. "Cookies" came about to enable timely retrieval of customized web pages.

Cookies can be used to identify web access by some user clients. Cookies are a general mechanism which server side Internet web connections such as, e.g., common gateway interface (CGI) scripts, can use to store and retrieve information on a client side of a hypertext transfer protocol (HTTP) connection. By using a persistent, client-side state, software file the interactive capabilities of web-based client/server applications has been increased. A cookie is a well-known term used for describing an opaque piece of software data held by an intermediary. A cookie is a holder of information. It cannot be used to get information off of a client's hard disk drive. Rather, a cookie can be used to save information entered voluntarily by a client and can be saved for future reference to avoid retyping of this information. Other example uses for cookies include, e.g., indicating a preference for viewing web pages in frames or text-only format, viewing a page in a particular language, storing a password and user name or other account number for sites that charge for viewing, and saving any other personal data needs which can be saved in a cookie so long as it isn't too long so as to exceed the 4K bytes limit for a cookie.

A cookie can be sent from an HTTP server to a client. Once sent, the cookie will be forwarded along with any request to the server from the client. HTTP servers are Internet servers which can contain hypertext software code such as, e.g., hypertext markup language (HTML). When a client on the Internet enters a universal resource locator (URL) address into a web browser, it is converted by a domain name server (DNS) into an IP address corresponding to a

file on a server. The HTML source code is sent from the server to the client's browser. The browser parses the code into several requests which can be sent to the server from the client. A server when returning an object to a client, can also send along a cookie, which the client can store on its workstation. Included in the state object is a description of the range of URLs for which that state is valid, i.e., the domain of the cookie. Any future requests made by the client which fall in that range, i.e. that domain, can include a transmittal of the current value of the state object from the client's browser to the server. It would be apparent to those skilled in the art, that references to HTTP requests, HTTP servers and HTTP clients in this document could also include other types of servers, clients and information transfer such as, e.g., data, media, audio, telephony, and streaming technologies.

For example, Netscape, Mosaic, and Microsoft Internet Explorer web browsers support cookie technology. Each cookie is a multipurpose Internet mail extension (MIME) header that can be used to exchange information automatically between a server and a browser without a user seeing what is being transmitted. The server can provide the user's browser a web page customized according to the pre-defined preferences contained in the cookie. Cookies can be used, e.g., by e-commerce shopping applications to store information about currently selected items, and for fee services to store registration information. Cookies can free the client from retyping a user ID on the next connection, and can store user preferences for the client such as, initial screens preferred upon entry to a domain. However, if a user disables the use of cookies, it can be difficult to identify access by a client user.

Thus a cookie can be useful in tracking a user's actions and preferences. For example, cookie data can be used to save values of data entered into a form. A cookie can be used by a web server site to store a user's preference information over several visits such as, e.g., how the user prefers to view the web page (in text or frame format), a user's name or address and preferred language. Conventionally designed cookies support only one domain, so a different cookie is needed for each domain. Unfortunately, this can require a large number of cookies to be placed on the user's hard drive. Once an architectural limit is reached, some cookies are also deleted. Also, if more than one person uses the same computer, unfortunately, no provision exists to alert the web server that a different user is accessing the site. If a user accesses a web site using a different computer, there is no provision to notify the web server of the identity of

the user. Also, if a user disables the use of cookies, then a domain has no way to identify the user requesting a web page view so as to permit customizing the view of the domain's web page. To some users, the automatic creation and retrieval of cookies raises privacy concerns. Browsers can allow users to disable the cookie feature, thus eliminating the tracking mechanisms.

Although useful, conventional cookies do not identify all attempted accesses to web sites. Using a proxy server presents special problems to those attempting to gather data regarding web access usage by users accessing the Internet via a proxy. It has proven especially difficult to track usage by such users which cannot be identified uniquely by a permanently assigned unique IP address. In addition, the proxy server does not forward all requests to the website (i.e., the server). Instead, the proxy server returns pages previously retrieved which it has stored in its cache.

Another tool exists to determine how many persons access web sites. Proxy servers, by returning cached pages rather than forwarding the request to the server, shield downstream users from the web servers they access. A document describing a methodology, "Basic Advertising Measures," is available at URL <http://www.fastinfo.org/measurement/pages/index.cgi/basicadmeasures>. The methodology can help a single site to determine how many persons saw a particular ad on a web site and clicked on the ad. Therefore, this methodology lets a web server know that someone came to the site, but it does not permit the web site to know who came to the site, unless they also use a cookie. The methodology is used for counting Internet banner ad impressions and clicks. The methodology was designed such that two compliant implementations would generate basic impression and basic click counts that differ by less than 5%. There are two basic methods for ad counting in use on the majority of the Internet today, i.e., ad requests (sometimes also known as ad insertions), and ad downloads. Ad requests refer to the method of counting an ad impression when a page containing the ad HTML is requested. The ad download method counts an ad impression when the ad media (in this case, an image) is requested from a server. The methodology defines an ad counter as a program that responds to browser requests (e.g., an image tag IMG SRC request, and an anchor tag A HREF request) related to advertising. A valid basic impression is counted only when the ad counter receives and responds to a request for an image from a browser. This image request must be the result of an IMG tag in the HTML page. In response, the ad counter returns a location redirect, specifying

the location of a file or other program that delivers the image media. A valid basic click is recorded only when an ad counter receives and responds to a click request from a browser. The click request is the result of a user clicking on an anchor tag in the HTML page. In response, the ad counter returns a location redirect, specifying the location of the destination for the ad. The methodology includes several mechanisms to defeat proxy caching. To defeat caching, the methodology requires the IMG SRC URL to be unique across page requests by a single browser. To ensure IMG SRC URL uniqueness, the methodology suggests inserting the current time with seconds, or a sufficiently large random number in the IMG SRC URL as the page is delivered to the browser. As would be apparent to a person skilled in the art, the methodology is rather complex and still only results in information including the number of ad impressions, with no identification of who accessed the ad, unless the user has enabled cookies. Thus, using a cookie is complementary to using the methodology to provide additional information to the basic ad impressions. A better approach is needed.

Another type of conventional cookie is a global profile cookie. A global profile cookie is provided by a global profile service. For example, a global profile service can provide ads to multiple web content providers. A global profile service can store a single file on a user's machine that includes identification information for that user. Different domains can then subscribe to the global profile service to permit the domains to use the global profile service to provide features such as targeted ad banners on the domains' web pages. The subscribed domains use the global profile service to perform broader analysis of user traffic across the subscribed domains using the global profile cookie. Unfortunately, if a domain does not subscribe to the global profile service, traffic by the user to the unsubscribed domain would not be tracked and/or analyzed.

Analysis of a client user can include tracking demographic attributes of the user. A demographic attribute can include a "pure demographic" attribute such as, e.g., a client user's age, gender, and salary range. Another demographic attribute can be obtained by analyzing the behavior of a client user.

Thus, what is needed is an improved method of identifying, tracking, analyzing and reporting Internet user access to web sites that overcomes limitations of conventional systems.

Summary of the Invention

A method, system, and computer program product for analyzing client user accesses to the Internet in a substantially real-time manner, in an exemplary embodiment, can include accessing raw data, processing the raw data using a core technology and interfacing the raw data with the core technology.

In an exemplary embodiment, interfacing can include accessing a proxy log including a proxy log data record having a field including a location requested by the client user, a first IP address of the client user making the request, an action requested by the client user, or a time of the request; accessing an IP address assignment log including an IP address assignment log data record having a field including a second IP address assigned to the client user, a userID of the client user, or a time window of assignment of the second IP address to the client user; and merging the proxy log and IP address assignment log to obtain the clean raw data including a virtual cookie identification data including a location, an action, or a userID.

An exemplary embodiment of the invention includes generating a virtual cookie. An exemplary embodiment includes identifying a user accessing the Internet via a proxy server, including accessing a proxy log, accessing an IP address assignment log, and merging the proxy log and the IP address assignment log to obtain virtual cookie identification data. In an embodiment, the method can be performed post-browsing. In another embodiment, the method can be performed real-time. In one embodiment the proxy server is owned, leased or operated by an Internet service provider (ISP). In another embodiment the proxy server is owned, leased or operated by a corporate network. In yet another embodiment, the proxy server is a caching technology or a logging technology that can observe and record activity of users. In an embodiment, the proxy log is a log of the caching technology or logging technology.

In an embodiment, the IP address assignment log is a dial-up log or a dynamically assigned IP address log. In another, the IP address assignment log is a statically assigned IP address log, where a network of workstations are assigned an IP address by a server.

In an embodiment, the proxy log can include a proxy log data record having fields including a location requested, a first IP address of the computer of the user making the request, an action requested, and a time of the request.

In another embodiment, the IP address assignment log includes an IP address assignment log data record having fields including a second IP address, a userID of the user being assigned the second IP address, and a time window of the assignment.

Another embodiment features virtual cookie identification data including a location, an action, and a userID.

In another embodiment, merging can further feature correlating the first IP address and the second IP address and the time of the request and the timewindow of the assignment to the user to determine the userID making the request. Although the IP address fields of the two log files can be referred to as a first IP address and a second IP address, respectively, the correlating step matches identical IP addresses and overlapping request timewindows to determine the user making the request. Information in other logs could also be correlated in this way.

Another embodiment outputs the virtual cookie identification data.

Yet another embodiment analyzes the virtual cookie identification data. In one embodiment, demographic analysis is performed using the virtual cookie identification data. In another embodiment, the analyzing step includes associating demographic information with the userID. Demographic information can include attribute information about the user, provided by the user. In yet another embodiment, psychographic analysis is performed using the virtual cookie identification data. Psychographic information can include attribute information about the user which is based on analysis of observed behavior of the user. In another embodiment, analyzing can include associating psychographic information with the userID. In another embodiment, analysis of the virtual cookie identification data is done post-browsing. In another embodiment, the analysis of the virtual cookie identification data is real-time.

In another embodiment, the raw data can be provided by a website, a store, or other provider which has access to user activity information.

In another embodiment, processing the raw data using the core technology can include receiving clean raw data from the interfacing step, processing the clean raw data using a raw data processor, processing output of the raw data processor to obtain an inventory-centric demographic hyper-cube, merging a plurality of the inventory-centric demographic hyper-cubes into a merged inventory-centric demographic hyper-cube, and analyzing the merged inventory-centric demographic hyper-cube.

In yet another embodiment, processing output of the raw data processor can include loading user demographics, loading user actions, detecting and removing user activity, determining behavioral interest groups, determining user profiles, and building the inventory-centric demographic hyper-cubes.

In another embodiment, loading user demographics can include accessing user demographics records including a userID of the client user, or demographic data of the client user; accessing a user demographics database; or adding the user demographic records to the user demographics database.

In another embodiment, loading user actions can include accessing user action records from the virtual cookie, including a userID of the client user, a location requested by the client user, or a userID of the client user; accessing a user action database, or adding the user action records to the user action database.

In yet another embodiment, detecting and removing atypical user activity, such as, e.g., that of robots, can include accessing a user action database, scanning records for an atypical client user such as a software robot or an administrative user, accessing the user demographics database, or removing the atypical client users from the clean raw data.

In an embodiment, determining behavioral interest groups can include accessing a user action database, accessing an interest group definition, matching the interest group definitions and the user actions to obtain an interest group record, accessing the user demographics database, or inserting the interest group records in the user demographics database.

In another embodiment determining user profiles can include accessing a user demographics database, accessing a profile definitions database, matching the user demographics database and the profile definitions database to obtain a user profile record, accessing the user profiles database, or updating the user profiles database with the user profile record.

In an embodiment building the inventory-centric demographic hyper cubes can include accessing user demographics, accessing user actions, or combining the user demographics and the user actions to obtain the inventory-centric demographic hyper-cubes including inventory-centric data hyper-cube files and a timestamp.

In another embodiment, merging of the hyper-cubes can include accessing a plurality of the inventory-centric demographic hyper-cubes, accessing a demographic date file, or merging

the inventory-centric demographic hyper-cubes and the demographic data file to obtain the merged inventory-centric demographic hyper-cubes.

The method of the present invention, in an illustrative embodiment can further include providing an interactive user interface to the core technology.

An exemplary embodiment of a system, method and computer program product of the interactive user interface can feature reporting Internet user behavior statistics on a per location basis. An exemplary embodiment can include a method for displaying location-specific reports as a user browses the Internet, including browsing the Internet using a browser, monitoring activity with the browser, observing a location browsed where the location includes content, requesting a report on the location, and displaying the report regarding the location.

In an exemplary embodiment, the method's content can be from a website.

In another embodiment, content is static or dynamically generated.

In yet another embodiment, monitoring activity can include using an activity monitor.

In yet another embodiment, requesting a report can include requesting the report from a report server.

In one embodiment, displaying the report can include displaying it on a report display.

In another embodiment, the browser is an Internet browser application program.

In an embodiment, the browsing step is performed by a user. In one embodiment, the user can be a producer researching audience for the location, an advertising sales person looking for a specific target audience, or an advertiser looking for a specific target audience.

In one embodiment, the activity monitor can perform steps including monitoring the browser using a separate browser window, monitoring the browser using a separate application or a separate applet, monitoring the browser using a plug-in module installed into the browser, or monitoring the browser with a module incorporated into the browser.

In another embodiment, report requesting can include requesting a demographic and behavioral breakdown of an audience of the location, requesting a targeted demographic and behavioral breakdown of an audience subset of the location, requesting historical traffic levels for the location, or requesting predicted future traffic availability for the location.

In one embodiment, the report server is running on a computer of a user, a separate computer from the computer of the user, or the computer of the user having an activity monitor integrated with the report server.

In another embodiment, the requesting of a location report can include sending the location, sending two or more preferences of a user, generating the report on a report server, or receiving the report from the report server. In yet another embodiment, the preferences of the user include a type of the report to be generated by the report server; or a display preference determining how the report is to be displayed. Further features and advantages of the invention, as well as the structure and operation of various embodiments of the invention, are described in detail below with reference to the accompanying drawings. In the drawings, like reference numbers generally indicate identical, functionally similar, and/or structurally similar elements. The drawing in which an element first appears is indicated by the leftmost digits in the corresponding reference number.

Brief Description of the Drawings

The foregoing and other features and advantages of the invention will be apparent from the following, more particular description of a preferred embodiment of the invention, as illustrated in the accompanying drawings.

FIG. 1A depicts a high level block diagram of an example implementation of the analysis technology of the present invention;

FIG. 1B depicts an example block diagram of a network illustrating client access to the Internet;

FIG. 2A depicts a flow diagram illustrating an exemplary implementation of generation of an exemplary virtual cookie according to the present invention;

FIG. 2B depicts a block diagram of an exemplary embodiment of a proxy server telecommunications network configuration;

FIG. 2C depicts a block diagram illustrating an exemplary embodiment of use of a conventional cookie;

FIG. 2D depicts a block diagram illustrating an exemplary embodiment of use of a conventional global profile cookie;

FIG. 2E depicts an example environment illustrating the virtual cookie and an example universal profile server of the present invention;

FIG. 3 depicts a detailed block diagram of an example embodiment of the present invention illustrating an example implementation of the core technology;

FIG. 4A depicts a flow diagram illustrating an example of loading user demographic data in an exemplary process database;

FIG. 4B depicts a flow diagram illustrating an example of loading user action data in an exemplary process database;

FIG. 4C depicts a flow diagram illustrating an example of detecting and removing activity of atypical users such as robots in an exemplary process database;

FIG. 4D depicts a flow diagram illustrating an example of determining user behavioral data in an exemplary process database;

FIG. 4E depicts a flow diagram illustrating an example of determining user profiles in an exemplary process database;

FIG. 4F depicts a flow diagram illustrating an example of building inventory-centric cubes in an exemplary process database;

FIG. 4G depicts a flow diagram illustrating an example process database technique of the present invention;

FIG. 5 depicts a flow diagram illustrating an example of cube validity merger in an exemplary data merger of the present invention; and

FIG. 6 depicts an exemplary computer system.

FIG. 7 depicts an exemplary embodiment of a traffic report display illustrating an example range of traffic summarized on an exemplary weekly basis;

FIG. 8 depicts an exemplary embodiment block diagram illustrating an example of interaction between an activity monitor, a report server and report display with an internet browser;

FIG. 9 depicts an exemplary embodiment of the report display;

FIG. 10 depicts an exemplary embodiment of the demographic report display; and

FIG. 11 depicts an exemplary embodiment of the targeted demographic report display.

Detailed Description of the Invention

The preferred embodiment of the invention is discussed in detail below. While specific implementations are discussed, it should be understood that this is done for illustration purposes only. A person skilled in the relevant art will recognize that other components and configurations may be used without parting from the spirit and scope of the invention.

An Exemplary Implementation of an Embodiment of the Invention

The present invention is directed to an improved demographic and behavioral analysis system architecture for use in, e.g., identifying tracking and understanding user behavior on the Internet and in traditional stores.

FIG. 1A depicts a high level block diagram of an example implementation of the analysis technology of the present invention. Specifically, FIG. 1A depicts a block diagram of an exemplary high level system architecture 100 according to the present invention. High level system architecture 100 includes, e.g., raw data 102, a raw data interface 104, a core technology 106, a user interface 108, and users 110. The raw data 102 is inputted into raw data interface 104. The output of raw data interface 104 is input into a core technology 106. Core technology 106 is described further with respect to FIG. 3, below. Core technology 106 is accessed interactively by users 110 via user interface 108.

Raw data 102 includes, e.g., Internet client user information including logs of web page views by web browsers of demographic profile information, and information regarding purchases by users. Raw data 102 can be made available by a website, an online service provider (OLSP), an internet services provider (ISP) or other entity tracking Internet client users, such as, e.g., a corporation, hereafter these entities will be collectively referred to as "customers," or users 110, although users 110 can be different than those providing raw data 102. Raw data files 102 can be in the form of database records and data files, such as, e.g., a proxy server log file, a web server log file, an ad (i.e., advertisement, such as, e.g., a banner ad) server log file, and user registration information. Raw data files 102 can also include, e.g., user purchase information, which could be obtained from a customer ISP or company, or, e.g., from the output of a super market grocery store's point-of-sale (POS) purchase card user purchase tracking file.

The format of raw data 102 could include log files and other files stored in a customer-specific data format. Raw data 102 files include information regarding “who, what, and where.” By analyzing raw data 102 files, high level system architecture 100 can be used to determine “who” on the Internet did “what” kind of action, and “where” did they do it. “Who,” can be the client user of the Internet such as those shown in FIG. 1B, below. “What,” can be the action the Internet client user performed. “Where,” can be the Internet location, e.g., universal resource locator (URL), or address of the domain and path where the Internet user performed the action, i.e., the web page requested. For example, a client user identified as User395, could have looked at (i.e. a page view action) a web page referred to as page94 (which could be a URL such as, e.g., <http://www.someplace.com/index.htm>).

Raw data interface 104 can be an application program that can be customer specific that processes raw data 102 files and sends the resulting processed output to core technology 106.

Raw data interface 104, in one embodiment, can read in raw data 102 files, can break down the files into useful data, and then can pass the data to a raw data processor 302, described further below with reference to FIG. 3.

Core technology 106 processes raw data to gain an understanding of inventory-centric demographics. Inventory-centric means tracking the demographics of the client user audience that visits each location, i.e., Internet location, such as, e.g., a server of HTTP data, audio, video, telephony, media, streaming technology, or other kind of data. Core technology 106 enables near real-time (or real-time) demographic reporting on a per location basis. Core technology 106 enables near real-time (or real-time) demographic reporting with drill-down analysis of specific target audiences on a per location basis. Core technology 106 enables near real-time (or real-time) searching for locations that best match a specified target audience. Core technology is described further with reference to FIG. 3 below.

User interface 108 can be used for searching and dynamically generating reports. User interface 108 enables users 110 to drill down through the data contained in core technology 106. For example, user interface 108 can permit analysis of demographic and behavioral data stored in core technology 106. User interface 108 can provide a product-specific user interface into core technology 106. User interface 108 can allow users 110 to interact with demographic and behavioral data. User interface 108 provides access to users 110 to access the functionality of

core technology 106. User interface 108 can be web-based in one embodiment. In another embodiment any other user interface could be used, such as, e.g., a client-server based interface.

Users 110 can include, e.g., any entity with a large amount of data, that wants to analyze the data. For example, large Internet server sites, ISPs, grocery stores, and corporations may desire to analyze large amounts of data tracking client user requests or purchasing. Users 110 can also include companies seeking to perform targeting promotions or advertising.

In one embodiment of the invention, Internet user behavior is analyzed. It will be apparent to those skilled in the art that the system can also be used with a traditional brick and mortar business environment. FIG. 1B illustrates an example environment. Referring now to FIG. 1B, the figure depicts an example block diagram of a network illustrating client access to the Internet. Specifically, FIG. 1 depicts a block diagram of an exemplary telecommunications network 120. Telecommunications network 120 includes a plurality of networks interconnected via the global Internet. An internet (with a lower case "i") is a network that connects multiple networks. The Internet (with a capitalized "I") is an internet which connects computer workstation hosts in many networks which communicate using the Internet protocol (IP). Each host of the Internet has its own IP address, which is used as a source or destination address in routing packets of information through the Internet.

FIG. 1B illustrates a variety of methods available for connecting to the Internet. For example, telecommunications network 120 includes a network 122 and a network 124 which are connected to Internet 158 via a proxy server 148. Specifically, network 122 is a token ring network including workstations 126, 128, 130 and 132. Network 124 is an ethernet network including workstations 136, 138, 140 and 142. The workstations in network 122 and the workstations in network 124 are connected to proxy server 148 via network connections as represented by lines 134 and 144. Lines 134 and 144 are logical connections and could represent a variety of different communications links and devices such as, e.g., cabling, gateways, bridges and routers. Proxy server 148 connects to Internet 158 via a connection to Internet 156. Connection to Internet 156 is a logical connection and could also represent a variety of different communications links and devices. Each of workstations 126-132 and 136-142 include a network interface card (NIC) for physically connecting to the other workstations on networks 122 and 124. It will be apparent to those skilled in the art that other network connections could

equally be used. Subscriber 146 is connected to proxy server 148 via a modem connection (such as, e.g., a dial-up connection) using modems 150 and 152 to proxy server 148. Proxy server 148 also serves to permit subscriber 146 to access Internet 158 even though it does not have a network interface card (NIC). The machine running proxy server 148 can also act as a network communications server (NCS) 242 (described further with reference to FIG. 2E. An NCS can provide network access to workstations so they can access the Internet by, e.g., a dial-up communications link. FIG. 2E described further below, illustrates an embodiment of the invention using a proxy server 148 and a separate NCS 242 for handling IP address assignment using an IP address assignment log 204. Subscriber 146 can dial into proxy server 148 using a modem 150. Subscriber 146 can be, e.g., a corporate user accessing a corporate network while out of town on business, a home user dialing up via a modem, such as, e.g., a cable modem connection or other means of access to the proxy server such as an integrated services digital network (ISDN) or a digital subscriber loop (DSL) or ISDN concentrator. It would be apparent to those skilled in the art that modems 150 and 152 in other embodiments could include any other dynamic access methods, such as, e.g., cable modems, digital subscriber line (DSL), or other means of remote or local access. It will be apparent to those skilled in the art that subscriber 146 need not be connected via a dial-up connection and could be coupled via, e.g., a leased line, a wireless connection, a dedicated link or other connection. In an exemplary embodiment of the invention, modems 150 and 152 are conventional analog modems that can operate at different speeds and can include various error-checking capabilities, and modulation protocols. An NCS manages a pool of IP addresses which it can assign to any of workstations 126-132, 136-142 and 146, to provide the workstations access to Internet 158. IP addresses can be assigned dynamically or statically, i.e., on a temporary or permanent basis, respectively, to users. In one embodiment, the server machine referred to as proxy server 148 can include the functionality of both a proxy server and an NCS. ISPs often route all their hypertext transport protocol (HTTP) traffic through a proxy server. Routing traffic through a proxy permits caching requests. Box 154 indicates that requests from the workstations it surrounds can be hidden from view of the remainder of the Internet 158. Proxy server 148 can act as a firewall to provide security to workstations on downstream networks 122 and 124.

In another embodiment, proxy server 148 is used by an entity other than an ISP, such as, e.g., a company with telecommuting employees dialing in via an NCS. In an embodiment, the NCS is a remote access device (RAD) which can be compliant with the dynamic host configuration protocol (DHCP). A RAD can be used to connect off-site users to a corporate network. These users can include, e.g., salespeople, and other business professionals who travel or telecommute rather than work in a fixed office location.

Internet 148 includes various networks connected together communicating via the Internet protocol (IP). Different networks can be coupled via a router. A router communicates between networks and is knowledgeable of workstations on multiple domains and can route information between those domains. For example, network 162 is coupled to Internet 148 via router 160 as indicated by line 172. Network 162 includes workstations 164, 166, 168 and 170 connected in an exemplary ethernet topology. Router 160 of network 120 routes IP packets from workstations on network 162 to other workstations on Internet 158. It is important to note that workstations 164, 166, 168 and 170 each have their own permanently assigned IP address.

By comparison, hidden workstations in box 154 can be assigned an IP address by an NCS and can have some hypertext transport protocol (HTTP) requests hidden by proxy server 128.

The hidden workstations can use a different IP address, i.e. a dynamically assigned one, each time they connect to the Internet 138. A network communication system (NCS) such as 242, below, can manage assigning a pool of IP addresses to the workstations it is responsible for, alternatively NCS functionality in the computer workstation of proxy server 148 could do so. It would be apparent to those skilled in the art that users 136-142 could also have permanently assigned IP addresses, known as statically assigned, assigned by proxy server 148.

Proxy server 148 running proxy server software can perform numerous proxy functions such as, e.g., caching of web pages. Caching of web page requests can save, e.g., as much as 50% of the traffic between connection to Internet 156 and Internet 148. Caching can be used to attain a high cache hit rate to decrease network traffic for an ISP and for enabling faster access time for users. Example cache protocols include, e.g., Internet cache protocol (ICP) and cache array routing protocol (CARP). Multiple proxies can be used to store large amounts of cached data. It will be apparent to those skilled in the art that when a proxy server is referred to in this document, the

proxy server could also be any other caching or logging technology that can observe and record user activity.

User activity can be obtained in a usable form from various data sites. In some cases, user activity information is readily available. In other cases, data can be processed into a usable form prior to analysis. For example, a virtual cookie can be used to take information from a proxy log file and can analyze the log file data and process it to prepare it for use as a raw data source. The virtual cookie thus is an optional, but not required, process of preparing user activity data for the present invention. FIG. 2A depicts a flow diagram illustrating an exemplary optional process which can be used to process file data for use as a raw data source. In one embodiment, the data processing step creates a virtual cookie. Specifically, FIG. 2A illustrates a more detailed block diagram of a raw data interface 104.

An example of a proxy server such as that used on ISP proxy server 128 is SQUID Internet Object Cache 2 available from FTP site squid.nlanr.net. SQUID-2 is derived from software developed and funded by the advanced research projects administration (ARPA) Harvest Project. SQUID-2 is a high-performance proxy caching server for web clients, supporting FTP, Gopher and HTTP data object requests. The SQUID-2 cache software is available only in source code, is relatively fast because it handles all requests in a single, non-blocking, I/O-driven process. SQUID-2 never needs to fork, is implemented with non-blocking input/output (I/O), keeps meta data and hot objects in virtual memory (VM), caches domain name server (DNS) lookups, supports non-blocking DNS lookups, and implements negative caching of failed requests. SQUID runs on all popular UNIX operating system platforms, such as, e.g., AIX, FreeBSD, HP-UX, IRIX, Linux, NeXTStep, OSF/1, Solaris, and SunOS, the OS/2 operating system platform, and the Windows/NT platform. A detailed description of SQUID is available at URL <http://squid.nlanr.net/Squid/>, and a frequently asked question (FAQ) list is available at URL <http://squid.nlanr.net/Squid/FAQ/FAQ.html>, the contents of which are hereby incorporated by reference in their entirety. Another example of a proxy server 128 is MICROSOFT Proxy Server 2.0 available from Microsoft Corporation of Redmond, WA.

In particular, FIG. 2A illustrates a flowchart 200 which depicts the process of creating a virtual cookie 214. Flowchart 200 includes as input, proxy logs 202a, 202b and 202c and IP address assignment logs 204a, 204b and 204c. Proxy logs 202a-c can reside on the same or

different proxy servers, or web servers of customers such as an ISP. IP address assignment logs 204a, 204b and 204c can also reside on the same or different proxy servers, or other servers.

Proxy logs 202a-c can contain requests from user client machines, such as, e.g., a subscriber. It will be apparent to those skilled in the art, that requests referred to as "HTTP requests," or "requests" could also include other types of requests from other types of servers by other kinds of clients, such as, e.g., data, media, audio, telephony, and streaming technology requests. The request can contain the requested URL (location), the IP address or number of the client user making the request, and the time that the request was made. Also, actions requested by a client user can also be captured which will usually be a web page pageview, although some users may perform another action, such as, e.g., may be selecting an advertisement or click-through, or other parsed HTML request, such as, e.g., an image request. Thus, in one embodiment, information may be logged relating to pageview actions, and in other embodiments, other actions can also be logged such as, e.g., click-through information to identify other behavior. Therefore, proxy logs 202a-c including location, action, IP address, and time of the requests, can be combined as indicated in processing step 206, and can be output as represented by line 210.

IP address assignment logs 204a-c can be maintained by, e.g., an NCS, an RAD or other type of server of, e.g., an ISP, a corporation or other customer entity. Internet client users can connect to the global Internet by using, e.g., a modem to establish a dial-up connection to, e.g., an ISP or customer. Once a client user is connected to the Internet via, e.g., an NCS of an ISP, the ISP can assign the user a temporary IP address or an IP number, which the client user can use for the duration of the user's connection to the Internet. IP address assignment logs 204a-c can also be consolidated as depicted in FIG.2A as part of raw data interface 104. IP address assignment logs 204a-c can include for a dial-up Internet access by a client user, the client user's user identification (userID), an IP address, and a time window during which the client user was connected to the Internet. In one embodiment, a userID is a unique identifier for a user. IP address assignment logs 204a-c including UserID, IP address, and time window logged on, can be combined as indicated in processing step 208, and output as represented by line 212 of raw data interface 104.

Virtual cookie 214 can take as input the output of steps 206 and 208 as indicated by lines 210 and 212, respectively. Line 210 can represent the output of processing of proxy logs 202a-c and line 212 can represent the output of processing of IP address assignment logs 204a-c. Virtual cookie 214 can merge the data contained in steps 206 and 208 and, can correlate using IP address (or IP number) and time to obtain the locations requested and actions requested by userID. Specifically, virtual cookie 214 can create a merged file which can include, e.g., for each location accessed, the action requested, and by what userID, which is indicated in step 216. By merging proxy logs 202 with IP address assignment logs 204, and by correlating records by time and IP address overlap, virtual cookie 214 can identify, e.g., all locations accessed by a specific user 222 (shown in FIG. 2E). Further demographic and psychographic analysis can be performed to create a profile for user 222 using the identified locations accessed by the user.

FIG. 2B depicts block diagram 248 which illustrates an exemplary network configuration for an example proxy server 148. Block diagram 248 includes at its base at the physical level dynamic access method 258, which in one embodiment could be, e.g., modem 132, an internal network interface card (NIC) 250 facing downstream networks 122 and 124, and an external network interface card (NIC) 252 which provides upstream access to Internet 158 via connection to Internet 156, which could be, for example, a router. An example router is a CISCO router available from CISCO Corporation of Mountain View, CA. Included in block diagram 248 are low level protocol drivers 254, transmission control program/Internet protocol (TCP/IP) network protocol stack 256, web proxy 260, IP address assignment log 204 and proxy log 207. In one embodiment, IP address assignment log 204 is a dial-up log, which could be, e.g., a log of dial-up subscribers 146 dialing up to access an ISP. In another embodiment, IP address assignment log 204 tracks static or dynamic assignments of IP addresses to users over time.

In one embodiment of the invention, IP address assignment log 204 and dynamic access method 258 run on a separate network communications server (NCS) computer than the proxy server, see FIG. 2B below. In this embodiment, proxy server 148 could include web proxy 260 and proxy log 202 and would have the proxy server software running on a separate computer from the NCS. It would be apparent to persons skilled in the art that other configurations could be used to implement a proxy server and a network communications server (NCS).

In another embodiment, an example NCS is a remote access device (RAD). The RAD can comply with dynamic host control protocol (DHCP). The RAD can be used to provide dynamic IP address assignment to network workstations connected through a proxy server. In this embodiment, rather than a dial-up log, a dynamic IP log tracks the assignment of IP addresses to the network workstations. IP addresses can be dynamically or statically assigned to the network workstations.

Proxy server 128 supports Internet access requests from downstream workstations 126-130, and 136-142, and subscriber 146. Requests can come into proxy server 128 through internal NIC 250 and can be handled by, e.g., web proxy 260, to open a connection to Internet 158 out through external NIC 252. Requests can also come into an NCS (shown in FIG. 1B as part of proxy server 148) via modem 152 from a subscriber 146 and can be similarly handled.

Proxy server software can perform logging functions. Each request from a workstation in box 135 to access Internet 158 is logged in proxy log 202. Proxy log 202 in a typical environment can log a location that a client attempts to request, e.g., a URL address. Proxy log 202 can also include a log of the action requested such as to open the URL, the IP address requesting the action, and the time at which the request was made by that IP address.

It would be apparent to those skilled in the art that proxy logs could also be the log from any caching or logging technology that can observe and record user activity.

A network communication server also performs logging functions. For example, when subscriber 146 attempts to log onto Internet 158 by initiating a connection via modem 150 to modem 152 of proxy server 148, an IP address assignment log 204 records information such as the time period that a subscriber was logged on or the time period an IP address was assigned (statically or dynamically) to a network workstation. Specifically, IP address assignment log 204 can track information about, e.g., subscriber 146, including, e.g., a user ID of subscriber 146, the IP address assigned to subscriber 146, and the time period logged on including, e.g., a start time along with either an end time or a duration. A dynamic IP address assignment log can record IP address assignments of a DHCP-compliant remote access device (RAD). Other alerts and logs can also be maintained.

FIG. 2C depicts a block diagram 262 illustrating the use of a persistent client state cookie. In block diagram 262, a user 264 accesses the Internet 158 through an client 304 having an IP

address. Block diagram 262 assumes that the IP address is either permanently assigned to client 266, such as, for workstations 164-170, or is temporarily assigned to client 266, such as, e.g., for workstations 126-132, 136-142 or subscriber 146, using an assigned IP address from proxy server 148.

In FIG. 2C, client 266 is connected to Internet 158 to access various servers such as, e.g., servers 268 and 270. If an administrator of server 268 wishes to be able to track accesses by user 264, a software tool, the cookie, has been developed to enable server 268 to do so. Specifically, if user 164 requests to view a particular URL (e.g. <http://www.something.com>), as illustrated by line 272, server 268 can then respond, as illustrated, by line 274 to client 266. The process of accessing a particular web page is now briefly described. When user 264 requests a website by entering a URL, the browser of user 264 parses the HTML source code which comprises the entered URL. Parsing involves breaking up the HTML source file into separate requests of the domain server corresponding to the URL. For example, the HTML source file could include several image tag references. An image tag reference (IMG SRC) can require the browser to request a graphical bitmap image for insertion in the hypertext document. Thus a request to view a URL on an server 268 or 270, can actually create several requests of the server. In response to these requests, server 306, for example, can send down to user 264, e.g., the requested text of the web page, and/or parsed images, associated with the URL requested in line 272. In addition, server 268 can send along an embedded software object, known as a persistent client state cookie 276, the "cookie." Cookie 276 can include a required name field which contains a value which may include information encoded within its value placed there by server 268 to identify user 264. In addition, cookie 276 can contain an expiration date and time, in Greenwich Mean Time (GMT), a domain of the cookie which is limited to a single domain, a path, and a security setting. Assuming user 164 is connected to Internet 158 via a connection, such as subscriber 146, then cookie 276 could be placed on the hard disk drive of the workstation of subscriber 146. The next time that user 164 dials in using subscriber 146 and attempts to access server 268, server 268 can also be sent by the browser the cookie 276 from the hard disk drive of dial-up subscriber 146 and can decode the information contained within cookie 276 in order to recognize user 264 and customize presentation of the webpage according to the preferences of user 264.

Second, there are issues associated with analysis of users caused by using proxy servers. By responding to a client's request with a cached version of a requested web page, the proxy server shields the client's request from the rest of the Internet. Efforts have been taken to get around the problems of tracking requests due to the use of proxy servers. Conventional attempts to analyze web usage for users, such as the methodology described in the background above, attempt to get around proxy servers but cannot track all user activity. The methodology uses a somewhat convoluted approach to track the number of users accessing an ad. If enabled, a cookie can be sent to the ad server as well. However, no information is provided as to what type of user accessed the ad if cookies are not enabled. Further, the approach described by the methodology fails to track usage when requests are cached, by proxy caching mechanisms. Popular ISPs, such as, e.g., AOL, use extensive caching to decrease overall network traffic. Thus, access to cached sites is not tracked by conventional ad tracking methodologies, since the requests are hidden behind the proxy server.

FIG. 2D depicts a block diagram 184 illustrating another way of tracking users on the Internet, in this case using a global cookie 289. The global 289 cookie is based on the idea that the more user actions observed of a user, the more information will be available about the user, and the more accurately can the user target by analysis. Block diagram 284 includes a user 286a accessing the Internet via an client 288 with a permanently assigned IP address. Block diagram 284 also includes a user 286b which is using the subscriber 146 work station as indicated within box 299 whose access to Internet 158 is via proxy server 148 which could be that of e.g., an Internet service provider (ISP). Specifically, user 286b is assigned an IP address by ISP proxy server 148. In block diagram 284, client 288 can be one of workstations 164-175 with a permanently assigned IP address. FIG. 2D illustrates how users 286a and 286b can be identified using a global cookie so as to provide better serving of advertisements, through the pooling of ad requests. Instead of using a separate cookie for each domain of, or even in addition to using separate domain cookies, e.g., servers 290 and 492, a single, global cookie 289 is used in this example. When servers 290 and 292 provide web pages to users 286a and 286b, they pool their banner ads by using an Ad server shown as global profile server 298. By pooling requests of, e.g., user 286a, a profile can be created based on previous historical access by user 286a. To implement the global profile server 298, a single global cookie 287 is placed on the workstation

of user 286a. Using the global profile cookie 287, global profile server 298 can analyze the browsing habits of user 286a and target an ad using the user profile 281 of user 286a. Servers 290 and 292 would need to have subscribed to the advertising services of global profile server 298 which can include user profiles 281 including ad preferences and viewing history for users 286a and 286b for web sites which have subscribed with global profile server 298. An example global profile server 298 is ProfileServer 4.0 available from Engage Technologies, Inc. of Andover, MA. Global profile server 298 only supports HTTP servers 406 and 408 which have subscribed to the global profile server 298 advertising services.

Assume user 286a attempts to access a web page on server 290. The browser of user 286a parses the HTML source of the page into multiple requests, such as, e.g., IMG SRC requests, as indicated by line 294. For example, a bitmap image can be sent down as indicated by line 296. In addition, an advertisement banner request can be parsed out which is then sent as a GET request to global profile server 298, including global cookie 287. In response to request 291, global cookie 287 can be used by global profile server 298, to access user profile 281 corresponding to user 286a to determine a banner ad to display in the requested web page. Once a banner ad is identified, global profile server 298 can send the banner ad to the browser of workstation client 288 of user 286a for viewing. Server 290 can then query global profile server 298 as indicated by line 291 and can receive results about user 286a as indicated by line 293 from a subscribed user profile _____. Global profile server 298 can store browsing information about user 286a on global profile server 298 for use in targeting future ads to user 286a. User profile 281 can include other information about user 286a such as, e.g., declared profiles and behavior profiles, for local browsing behavior and web wide (so long as a subscribed server 290, 292). Declared profiles would need some how to be captured, e.g., during access to an ad, user 286a would need to offer information, or such information would need to be supplied to subscribed server 290 and 292, which would need to capture and forward such information to global profile server 298.

Now suppose user 286a requests a web page to be opened from the domain of server 292, as indicated by line 285. server 282, if it has subscribed to global profile server 298, will include a parseable request to global profile server 298, similar to that described with reference to server 290. Global profile server 298 would be sent the same global cookie 287 by the browser of user

286a, as indicated by line 295. Global profile server 298 would send a targeted ad as indicated by line 297 to the browser of client 298 of user 286a. Global profile server 298 would pool the information gleaned regarding user 286a from the multiple subscribing servers 290 and 292. For example, if other users which had accessed server 292, had also requested a URL from server 292, global profile server 298 could place an ad on the requested web page from Server 292 to user 286a, to direct user 286a to services on server 290. Thus, global profile server 298 can pool behavior from multiple subscribed server 290 and 292 sites to more narrowly target advertising to user 286a, based on user profile 281.

Therefore, using global profile server 298, multiple subscribed servers 290 and 292 can benefit from advertisement serving using a single global cookie 287, if the servers 290 and 292 subscribe to the global profile service 298. Note that only one 287 cookie was necessary to be placed on client 288 for serving advertisements for multiple servers 290 and 292 domains. servers 290 and 292 are not provided the contents of global profile cookie 287. Servers 290 and 292 outsource their advertising to global profile server 298. In order to use global profile cookie 287, servers 290 and 292 need to subscribe to the services of global profile server 298, which itself is sent the global profile cookie 287. For global profile server 298 to be able to serve an ad banner to a user, the user must have enabled the use of cookies.

If user 286b attempts to access a website on a subscribed server, Server 290 and server 292, when connected via subscriber workstation 146 and proxy server 148, the browser of user 286b could similarly parse the requested web page and then could request part of the HTML from servers 290 and 292, and could similarly request ad banners from global profile server 298 by sending global cookie 289 to global profile server 298 to identify user 286b.

Global profile cookies 289 and 287, provide the advantage of using a single global profile cookie for permitting observation of behavior of users across multiple subscribed domains. However, using global profile cookies has limitations. For example, servers 290 and 292 must be subscribed with global profile server 298. Use of a global cookie still requires that the cookies feature of a user's browser be enabled. If servers 290, 292 do not subscribe to global profile cookie 287, 289, then browsing of non-subscribed sites by users 286a and 286b is not tracked. An example of an ad server of this sort, is, e.g., <http://www.doubleclick.com>.

Thus, global profile server 298 can observe behavior of anonymous visitors across multiple subscribed web sites. However, browsing habits to non-subscribed web sites are not captured by global profile server 298. Global profile server 298 can build an interest profile for users 286a and 286b based on which subscribed sites global profile server 298 observes users 286a and 286b browsing.

FIG. 2E depicts an example environment illustrating the virtual cookie and an example universal profile server of the present invention. Specifically, FIG. 2E depicts a block diagram 220 illustrating the use of a universal profile server 240 according to the present invention. Universal profile server 240 can use post browsing analysis to create a virtual cookie 214 to track web browsing behavior by users, in one embodiment of the invention. In another embodiment of the invention, universal profile server 240 can create the virtual cookie 214 in real-time. Virtual cookie 214 advantageously does not actually require that any user enable the cookie feature of browsers. Virtual cookie 214 also provides much more targeted information regarding user browsing habits than available through any conventional behavior tracking approaches. Instead of only providing measurements of the number of users to a particular site, virtual cookie 214 can provide much more robust analysis information regarding not only how many visited sites, but also, e.g., what types of client users, i.e., who visited a location, what action was performed at the location visited, and where was the location visited.

The present invention uses the IP address of the workstation of users 222a and 222b in order to uniquely identify users 222a and 222b. It is conventionally thought that the IP address of the workstations of users 222a and 222b is insufficient to track all users, because of the large amount of dynamic IP allocation. As depicted in FIG. 2E, user 222a has a permanently assigned IP address, assigned by network communications server (NCS) 242. In an alternative embodiment, NCS functionality is contained on the same machine as the proxy server software, e.g., proxy server 148. User 222b accesses Internet 148 using a temporarily or dynamically assigned IP address from NCS 242, so user 222b can use a different IP address each time it accesses Internet 158. Worse still, since many different workstations 126-122 and 136-142 can also use the same IP address as subscriber workstation 146 of user 222b, HTTP servers 226 and

228 (or other data, media, audio, video, telephony, or streaming technology servers) can never definitively know whether a particular user 222b is accessing servers 226 and 228.

Thus, user 222b conventionally can not be uniquely identified by an IP address. However, using, e.g., a post browsing analysis technique (or a realtime technique in an alternative embodiment) of the present invention, named a virtual cookie 214 (recall FIG. 2A above), all web site browsing of user 222b and 222a can be tracked and analyzed by individual user. The reader should appreciate that although the inventors have named the analysis tool "virtual cookie," it is not in fact a cookie at all, and does not require enablement of browser cookie features. According to one embodiment of the present invention, universal profile server 240 uses a virtual cookie 214 to identify and analyze all web sites browsed by users 222a and 222b. Using virtual cookies 214a or 214b, all sites accessed by the users can be analyzed with no cookie needing to be stored on HTTP client 224 or workstation 146 of users 222a and 222b, respectively.

Therefore, all locations requested by all user web browsing activity can be tracked and analyzed without the need for placing a cookie on a user's workstation, and this virtual cookie works across all websites, not only a subscribed subset of web server locations as provided by a global cookie.

Further, virtual cookie 214 enables tracking and analysis of requests which were completed by the proxy server as a result of cache hits. In the case of a user 222a on an HTTP client 224 (or other data, media, audio, video, telephony or streaming technology client) such as one of workstations 164-170, 126-132, or 136-142 accessing web pages of HTTP servers 226 and 228 (or other servers) using a permanently assigned IP address through proxy server 148, all traffic can be tracked by virtual cookie 214a. Conventionally, HTTP client 224 would request websites from HTTP servers 226 and 228, as represented by lines 230-232 and 236-238, but would often rather receive a cached version of the requested web pages from the proxy server 148 as represented by lines 244 and 246. Thus, the proxy server would shield requests 230-232 and 236-238 from analysis detection on the rest of the Internet. Virtual cookie 214 tracks all requests by HTTP client 224 (and subscriber 146), including those for which a proxy server returned a cached web page to the client. In the case of user 222a, using a permanently assigned IP address, the user's browser is configured to use the proxy server. If HTTP client 224 is assigned an

address by a DHCP compliant RAD type NCS 242 (as shown), then the IP address assignment log 204 can track all locations accessed by user 222a.

If HTTP client 222b is a dynamically assigned IP address device, assigned via an NCS 242, then the client can also be analyzed and tracked in the same way as described with reference to statically assigned IP address devices. Specifically, In the case of a user 222b accessing HTTP servers 226 and 228 using a temporarily assigned IP address, it was conventionally thought difficult or impossible to track all traffic. Virtual cookie 214 can be created by using analysis after browsing by user 222b is completed and logged. This post browsing analysis can be performed at the proxy server as illustrated by virtual cookie 214a of FIG. 2E. Alternatively, post browsing analysis can be performed at a separate server such as, e.g., universal profile server 240 with access to the log data necessary for creating virtual cookie 214b.

A significant advantage of universal profile server 240 over any conventional profile server technology is its ability to identify users accessing Internet 158 via a proxy server 148, such as user 222b using subscriber workstation 146. A very large portion of the Internet population accesses the Internet via proxy servers, and in particular via proxy servers of internet service providers (ISPs) and other corporate entities. For example, ISP American On Line (AOL) has on the order of 15 million users whose IP addresses can vary each time they access Internet 158. If a large portion of users disable the use of cookies, there is no way to accurately identify access to HTTP servers 226 and 228 by users 222b, for example. Often all HTTP (and data, media, audio, video, telephony and streaming technology) traffic is sent through proxies to take advantage of proxy functions such as caching. With access via a proxy server, caching of web pages conventionally prevents accurate tracking of web site requests. Using the universal profile server 240 also provides the advantage of enabling cross web site analysis. For example, a web client may go to a combination of web sites which when analyzed together can indicate a particular attribute about the user.

Creating a virtual cookie 214 as already described above with reference to FIG. 2A, permits tracking and analyzing all browsing activity of users 222a and 222b. The technique maps an IP address to a userID. In one embodiment, this mapping is performed post-browsing. In another embodiment, this mapping is performed in substantially real-time. In one embodiment, virtual cookie 214a can be determined on proxy server 148, which can, e.g., be a proxy server.

In another embodiment, virtual cookie 214b can be created on a separate universal profile server 240. Universal profile server 240 can be a separate server computer or several computers with connectivity to Internet 158.

Virtual cookie 214a can identify user 222b by using information contained on proxy server 148 including requests to servers 226 and 228, and contained in logs on NCS 522. Conventionally, users 222b could not be identified by IP address, since it could have changed with every access. Thus, specific browsing habits of such users were only accessible by using a conventional cookie and without one, only general information was available. Given proxy logs 202 from, e.g., an ISP, browsing of users 222b of the ISP could only be reviewed generally because the proxy logs 202 only contain the requesting IP address, which does not uniquely identify a specific user 222b. Instead, the IP address represents a number of different users 222b, since a pool of IP addresses are assigned to a variety of users 222b by NCS server 242, at the time of IP address assignment.

Virtual cookie 214 can be used to analyze demographic and behavioral information about client users of the Internet 158. User activity data from, e.g., a virtual cookie, or user activity recorded by websites, stores, or other entities can be used and analyzed.

Demographic information can include, e.g., attribute information about a given user that is provided by the given user. Demographic information can be collected by an ISP or website, for example. Demographic information is often misleading. For example, an Internet user can often attempt to protect his or her privacy by withholding information or providing intentionally false or misleading information in a profile request of, e.g., the ISP. Demographic information can be collected from registration and from other sources.

Behavioral information can often be substantially different from demographic information attributes provided by a given user. Behavioral information includes observed behavior based attributes for a given user. Behavioral information is based on tracking observed behavior to create a behavioral profile for a given user. Since behavioral information tracks real user behavior, it is thought to be often more trustworthy than entered or claimed attribute information.

Using an optional (but not required) virtual cookie for processing user activity into a raw data source, a universal profile server 240 can associate user demographic information with an

identified user. For example, once a userID is determined for a client user, demographic information can be associated with the userID. For example, demographic information can be obtained by an ISP during the registration process. Demographic information is also often entered into servers during interaction between the user client and the server. However, demographic information captured by servers is often difficult to easily access and can be retained as proprietary by a given server.

By tracking all web site browsing activity by a client user, universal profile server 240 can prepare a behavioral profile based on observed behavior of users. By using the virtual cookie of the invention, this behavior information is more easily accessed and can be used as a highly reliable proxy for less reliable, less easily accessible demographic information.

Demographic information which can be collected about users and observed behavioral information compiled from the virtual cookie can then be analyzed in combination to provide, e.g., targeted advertising, targeted e-commerce offerings, and customized or personalized content, products and services. Analysis can be performed post-browsing or in real-time.

Referring to FIGs. 2A - 2E, universal profile server 240 can store the information output from flowchart 200 in step 210 and can perform further analysis on the information, using the information in the virtual cookie 214 as an index of information regarding browsing habits of users 222a and 222b. Additional profile information could be collected and associated with a given user 222a and 222b by associating the information with the userID, in universal profile server 240.

For example, universal profile server 240 could track user demographic information. A demographic information profile can be gathered about a user. A user's demographic profile information can include information such as, e.g., user ID, nicknames, aliases, e-mail addresses, home post office addresses, home city and state, home zip codes, work post office addresses, work city and state, work zip codes, home telephone numbers with area codes, work telephone numbers with area codes, home and work fax numbers, personal URL homepages, favorite URLs, preferred languages, and other user demographic information. User demographic information for a given user can also include, e.g., gender, age, national origin, race, orientation, marital status, weight, height, other dimensions, music preferences, drinking preference, smoking preference, education attained, income brackets, occupation, years employed, particular interest

groups such as, e.g., golf, fishing, sewing, safety, women's issues, and for business accounts, other interest groups such as, e.g., industry areas, employer information, size of the business, sales of the business, earnings of the business, number of employees employed by the business, the business type, SIC code, and industry SIC code, business location, business size (small, medium, large, multi-national business), other company information such as, company e-mail information, address information, telephone, fax, and company home page URL.

In addition, user psychographics, or behavior, can be observed and associated with a userID of a user to provide perhaps an even more accurate profile of the user. Behavioral information is tracked based on analyzing the virtual cookie 214 for a user including the locations browsed and actions taken by the user. The sites accessed by users 222a and 222b would be generated by virtual cookie 214. Based on the sites visited by a given user, universal profile server 240 can place a user in one or more categories. For example, if a user frequents many golf club manufactures' sites and golf course condition sites, then the user might be placed in a golfing enthusiasts' interest group. If the user visits many travel related sites such as, e.g., sites regarding remote vacation destinations, or cruise itineraries, the user might be placed in a travel enthusiasts' interest group. Finally users which frequent sites associated with luxury cars, golf and international travel sites, might be placed in an upper income focus category. Thus behavioral analysis could be used alone or along with demographics to analyze users. User psychographics, or behavioral information can indicate browsing habits which might conflict with designated profile information and demographic information. For example, a particular person might indicate that they are of a particular income bracket, but based on buying patterns or web locations accessed, as compared to other users, it might be determined that the user could be of a higher, or lower income bracket than declared. Thus, psychographics or behavioral analysis can indicate an expected demographic profile based on behavior analysis of comparable users 222a and 222b. Analysis decisions could be made such as, e.g., determining whether to trust a user's declared profile, or whether to rather rely on the behavior observed in virtual cookie 514. User behavioral information could include the history of URLs visited, and advertising banners selected. Interest group profile categories could be created based on certain observed behavior as recognized by analyzing browsing history captured in virtual cookie 214 of users 222b.

FIG. 3 depicts a detailed block diagram 300 of an example embodiment of the present invention illustrating an example implementation of the core technology. Specifically, FIG. 3 includes a detailed description of core technology 106. Block diagram 300 details components of core technology 106, including a raw data processor 302, a process database (DB) 304, a data merger 306, and a data analyzer 308.

Raw data processor 302 interacts with the raw data stream sent by raw data interface 104. Due to the large volume of raw data that potentially must be handled, speed and memory efficiency are critical. For example, an Internet portal site can easily have 10 gigabytes (GB) of raw data to process each day. Raw data processor 302 efficiently handles incoming raw data, identifying known users, actions, and locations, and prepares the raw data file records for input into process database 304. Raw data processor 302 takes as input the output of raw data interface 104. Raw data processor 302 can be thought of as manipulating raw data 102 and cleaning up the data for processing by the process database 304. Raw data interface 104 is described further with respect to FIG. 2A, above.

Process database 304 performs in-depth analysis on the processed data. Individual behavioral demographics can be generated based on client user activity. Inventory-centric demographics can also be calculated by process database 304. Process database 304 converts the raw data into a cleaned up form and generates an inventory-centric demographic hyper-cube. The inventory-centric demographic hyper-cube (ICDHC) or "cube" holds demographic information by location. An inventory-centric demographics hyper-cube is an n-dimensional cube of data including, for each location, demographic information including, e.g., pure demographic information (such as, e.g., age, sex, and occupation), and behavioral demographics, (i.e., generated by observing behavior of a client user, such as the locations requested on the Internet by the user). Process database 304 is described further below with reference to FIGs. 3, 4A, 4B, 4C, 4D, 4E, 4F, and 4G.

Data merger 306 can support processing of data which has become too large for an individual computer. For example, data merger 306 can take fully processed data from multiple process databases 304 and can combine the data into a single, consolidated data set. Specifically, data merger 306 can merge multiple cubes contained in process database 304 to form a single consolidated ICDHC cube. Data merger 306 enables merger of data from, e.g., several days,

weeks or months. Data merger 306 also enables merging of data from a large data set that is too large to be easily processed. The large data section set can be split into subsets and can then be processed separately. Data merger 306 can merge a plurality of ICDHC cubes by, e.g., averaging, running rolling averages, and averaging with data from a former year to detect shifts from previous years. Data merger 306 permits merger of such data without requiring processing of all data at once, e.g., to process 6 months of data, one need not load all 6 months worth of data, but can rather load, analyze and process each of the 6 months separately and then can merge the results to obtain a merged ICDHC. Thereafter, the merged ICDHC can be stored rather than storing the data of all 6 months. It will be apparent to those skilled in the art that several advantages are obtained from processing a plurality of cubes and then merging the separately processed results. Data merger 306 is described further with reference to FIG. 5 below.

Data analyzer 308 can enable near real-time (or real-time) in-depth reporting and search capabilities on the processed data set. Queries can be made against the ICDHC cube. Queries on, e.g., individual locations, target audience versus individual locations, and searches based on target audience, are supported. Data analyzer 308 enables near real-time (or real-time) demographic reporting on a per location basis by accessing a cube. Data analyzer 308 enables near real-time (or real-time) demographic reporting with, e.g., drill-down capability on specific target audiences on a per location basis by accessing a cube. Data analyzer 308 enables near real-time (or real-time) searching for locations that best match a specific target audience by accessing a cube. Data analyzer 308 allows users 110 to query the demographics for a particular location.

For example, user 110 can query the breakdown by sex of client users accessing a location, i.e., e.g., a page on a web site on the Internet. Then user 110 can drill down within the percentage of females accessing a site, to determine, e.g. what percentage of the females are interested in sports as demonstrated by their behavior.

In another example query, a user 110 can seek a target audience. For example, user 110 can seek a target audience of male, ages 2-17, and interested in yo-yos. A query with such information could yield Internet locations most likely to yield the targeted male 2-17 yo-yo enthusiast audience. Then user 110 could use this information to target an advertisement directly to those client users by placing ads on the resulting locations.

The process database 304 component of core technology 304 is now described with reference to FIGs. 4A-4G.

FIG. 4A depicts a flow diagram 304A illustrating an example of loading user demographic data in an exemplary process database 304. Flow diagram 304A illustrates how user-specific demographic information that raw data processor 302 has collected can be stored in process database 304. Specifically, flow diagram 304A depicts demographic data can be indexed by userID as represented by processing step 402. The user demographic data in processing step 402 is then added as indicated by line 404 (representing adding records) to user demographic data 406 of process database 304. An example of user demographic data for client users follows. For a user user28, demographic data can include, e.g., gender is male, age is 18-21, and occupation is student. For a user user29, demographic data can include, e.g., gender is female, age is 2-17, and occupation is student. The data records for users user28 and user29 can both be added to process database 304 as illustrated in FIG. 4A. Such demographic data can be obtained from, e.g., a registration process.

FIG. 4B depicts a flow diagram 304B illustrating an example of loading user action data in an exemplary process database 304. Flow diagram 304C represents an example technique by which user action records collected by raw data processor 302 can be stored in process database 304. Specifically, flow diagram 304B depicts user action data can be indexed by userID, including location accessed (e.g., the URL requested), and action requested (e.g., a page view, or click through) as represented by processing step 408. The user action data in processing step 408 is then added as indicated by line 410 (representing adding records) to user action data 412 of process database 304. An example of user action data for client users follows. A first record of user action data can include, e.g., location is URL requested, <http://www.somewhere.com/sports/football.htm>, action is pageview, requested by user, user28. A second record of user action data can include, e.g., location requested is ad934 clicked on webpage <http://www.somewhere.com/sports/football.htm>, action is clickthrough, requested by user user28. A third record of user action data can include, e.g., location is product10035, action is purchased by user user28. The data records for the three user actions performed by user user29 can be added to process database 304 as illustrated in FIG. 4B. It would be apparent to those skilled in the art that other actions can be added such as, e.g., requests for streaming media.

FIG. 4C depicts a flow diagram 304C illustrating an example of detecting and removing robots in an exemplary process database 304. Some raw data 102 that was collected can contain non-representative data. For example, user actions requested by in-house system administrators monitoring a web site location, and requests from visits by computer robots can be logged as activity, but such requests are not really typical client user requests of the sort that are sought to be tracked. By analyzing actions, atypical user data activity can be detected and removed in order to yield more accurate resultant data. Specifically, flow diagram 304C depicts detection and removal of robot requests from user demographics 406 of process database 304, by detecting actions by robots by scanning action records 412. Flow diagram 304C includes scanning records represented by line 414 of actions database 412. Scanning records step 414 finds requests by robots or spiders, i.e., computer software agents creating by parsing routines and search engines, for example by using statistical methods. From step 414, step 416 can be performed. In step 416, robot action records can be removed as represented by line 418, from user demographics database 406 of process database 304, by removing users identified as non-representative of client users. It would be apparent to those skilled in the art that other means could be used to remove atypical users from further processing.

FIG. 4D depicts a flow diagram 304D illustrating an example of a technique for determining user behavioral data in an exemplary process database 304. By analyzing user action, behavior-based demographics can be determined. For example, since user28 visited the page known as <http://www.somewhere.com/sports/football.htm> one can assume that user28 is interested in football. Specifically, flow diagram 304D depicts how an interest group definitions database 420 can be matched with user actions database 412, as indicated by line 422, and can be processed as shown by an interest group builder 424 step. Interest group definitions 420 can include actions that a client user performs to be considered part of that particular interest group. For example, visiting the <http://www.somewhere.com/sports/football.htm> page could be the action that triggers inclusion in the football interest group. Interest group builder 424 can match each user's actions to the interest group definitions 420 and then can add additional behavior-based demographics to user demographics 406. Interest group builder 424 can then insert records as shown by line 426 into user demographics database 406 of process database 304.

FIG. 4E depicts a flow diagram 304E illustrating an example of determining user profiles in an exemplary process database 304. Profile definitions 428 are sets of demographics that a client user has in order to be considered part a particular profile. For example, in one embodiment the audience of client users can be divided into separate buckets for separate analysis. Specifically, flow diagram 304E depicts user demographics 406 and profile definitions 428 can be matched as indicated by line 430 as part of a profile determination step 432. As indicated by line 434, records of a user profiles database 436 can then be updated using the output of profile determination 432.

An example of profile determination of client users follows. In one example, males and females may be analyzed separately. In another example, client users of different age group ranges could be analyzed separately. A third example could analyze groups divided up by sex and age into separate buckets, to obtain for example, separate analysis for males 2-17 and males 18-21, and females 2-17 and females 18-21. The resulting data records can update records, as shown in line step 434, of user profiles database 436 of process database 304, as illustrated in FIG. 4E. User profiles 436 gives, e.g., several buckets of information for a given location, i.e., additional differentiation can be provided because data is stored in separate buckets. By storing demographic data in various granular buckets, additional drill-down analysis is enabled by users 110 analyzing the resultant data. Another embodiment could use standard clustering techniques on a site or per location basis to better separate users into groups.

FIG. 4F depicts a flow diagram 304F illustrating an example technique of building inventory-centric cubes, such as, e.g., inventory-centric demographic hyper-cubes ICDHC, in an exemplary process database 304. Flow diagram 304F illustrates how a profile builder 440 can interact with process database 304 to split out a plurality of files, collectively referred to as a cube 450. Process builder 440 generates inventory-centric demographic hyper-cubes including data on multiple locations, and for each location tracks information such as, e.g., the profiles of the people in the various buckets such as the average demographics, and a timestamp indicating what data set the cube was generated from, i.e., including the effective date of the data. Specifically, flow diagram 304F depicts user demographics database 406 and user actions database 412 being matched, as indicated by line 438, by profile builder 440. Profile builder 440 can convert information in the database into a cube 450. Profile builder 440 can combine records at the same

location with the same profile and output a cube 450. The cube 450 can include the averaged demographic data for each of the location's profile types. The timestamp file 448 can contain the timestamp of the data set that the cube 450 was generated from.

FIG. 4G depicts a flow diagram 460 illustrating an example process database 304 processing technique. Flow diagram 460 begins with step 462 and can continue immediately with step 464.

In step 464, process database 304 can load user demographics as described further already with respect to FIG. 4A, above. From step 464, flow diagram 460 can continue with step 466. In step 466, process database 304 can load user actions as described further already with respect to FIG. 4B, above. From step 466, flow diagram 460 can continue with step 468.

In an alternative embodiment, steps 464 and 466 can be performed in parallel. It should be appreciated that the order of the steps of the process can be varied within the spirit of the invention, as would be apparent to those skilled in the art, so long as any data required as an input to a process step is available at the time of performance of the given step, i.e., so long as there are no time dependencies requiring output of a particular step to be used as an input to the other step.

In step 468, process database 304 can detect atypical users, such as, e.g., a robot, and can remove their actions as described further already with respect to FIG. 4C, above. From step 468, flow diagram 460 can continue with step 470.

In step 470, process database 304 can determine interest groups including, e.g., behavioral analysis as described further already with respect to FIG. 4D, above. From step 470, flow diagram 460 can continue with step 472.

In step 472, process database 304 can determine user profiles as described further already with respect to FIG. 4E, above. From step 472, flow diagram 460 can continue with step 474.

In step 474, process database 304 can build inventory-centric cubes as described further already with respect to FIG. 4F, above. From step 474, flow diagram 460 can end with step 476.

The data merger 306 component of core technology 304 is now described in detail with reference to FIG. 5.

FIG. 5 depicts a flow diagram of data merger 306 illustrating an example of cube merger in an exemplary data merger of the present invention. The example hyper-cube merger includes

a demographic validity merger in the example embodiment. Generally, data merger 306 enables merging a plurality of hyper-cubes as described briefly above with reference to FIG. 3.

Specifically, FIG. 5 depicts an exemplary flow diagram 500 of exemplary data merger 306 including, e.g., a plurality of ICDHC cubes 450 (including, e.g., cubes 450a, 450b, 450c, 450d, 450e, and 450n), and a file of demographic dates 502. Cubes 450 a-n could be generated from a plurality of different data sets. Data sets of raw data 102 can usually include captured log data from, e.g., different days, weeks, months, or years. The plurality of data sets could also be the result of dividing a large data set into multiple process databases 304 by splitting the large data set into several smaller ones. A “demographic” can mean a particular demographic attribute (i.e., pure demographic or behavioral information) about a given user. Demographic dates 502 can include a file containing a date for each demographic that specifies when that demographic was first valid. As demographics are deleted or changed, e.g., the values for the deleted or changed demographic may no longer be valid. Tracking the validity of each demographic is important to maintain data integrity. For each demographic, demographic dates file 502, a file including the date that the demographic was first valid can be maintained. Thus, some data may no longer be valid such as, e.g., where a demographic was changed or was deleted. It should be apparent to those skilled in the art, that demographic dates 502, could also be, e.g., behavioral dates, such as, e.g., interest group behavioral attributes, or other data stored about client users.

Flow diagram 500 includes merging the plurality of cubes 450 along with demographic dates 502 as represented by line 504 and a profile merging process in step 506 to obtain merged ICDHC cubes 510a, 510b, and 510c.

In profile merger step 506, for each location, for each profile type, the portions of the profile that are still valid are merged. Profile merger step 506 of the exemplary flow diagram of FIG. 5 of data merger 306, can read in the different cubes 450a-n and builds a new merged cube 510a-510c based on the demographic information in cubes 450a-n. Profile merger 506 goes through each location and finds so-called “buckets” that are the same and then averages the values of the buckets. Averaging/merging of demographic bucket values can eliminate daily fluctuations in usage of locations. In one embodiment of the invention, merging can be done by averaging the demographics from each of the valid cubes. In an alternative embodiment, rolling averages or averaging in the previous year’s data, e.g., can be done to take into account

seasonality differences in data. Profile merger 506 can account for the fact that not all profiles may be up to date by merging only valid portions. For example, if a demographic is added today, then profile merger 506 will not merge in yesterday's invalid values of the demographic. Profile merger 506 of data merger 306 permits creation of, e.g., multi-day, multi-week, multi-month, and multi-year cubes and permits analysis to be performed on a combination of the cubes. Merged cubes 510a-c is an inventory-centric demographic hyper-cube that includes the location-specific demographic information generated by combining the demographics from cubes 450a-n.

In one embodiment, a requirement of profile merger step 506 is that the cubes to be merged must have been generated in a consistent manner. While each location can use a different clustering of users, the particular clustering that each location uses must be used during the generation of all of the ICDHC cubes being merged. Consistent clusters enable the profiles to be merged accurately since a profile from one cube can be combined with other profiles that have the same cluster membership. For example, if for a certain location we split the users by sex, male and female, we would generate two profiles where the demographics were averaged for males in one profile, and for females in the other profile. When we later generated a second ICDHC, we would get another pair of profiles, one each for male and female. Since these were generated consistently, these two sets of profiles can be combined accurately. Simple mathematical analysis can show that the merged set of profiles generated by combining any number of consistent profile sets will be substantially accurate as one profile set generated using the raw data used to generate the ICDHC being merged. On the other hand, if the ICDHC being merged have one ICDHC clustered by sex and the other ICDHC clustered by age, there is no meaningful method of merging these to form a single ICDHC.

Merged cube 510a-c can then be analyzed by users 110 using user interface 108 in conjunction with data analyzer 308 of core technology 106. Using data analyzer 308 and user interface 108, users 110 can perform in-depth analysis in a rapid manner. Conventionally, log file analysis provided only shallow, single level data, with limited searching capability and no ability to drill down. Conventional data mining on the other hand enabled some in-depth analysis, but requires extensive time and costly processing, e.g., queries taking several minutes and performing analysis on high performance, expensive super computer machines might be needed. Using the present invention, on the other hand, inexpensive, in-depth analysis can be

performed in a near real-time (or even real-time) manner. For comparison, it might in one embodiment only take 10 seconds to generate a drill-down analysis, permitting accessing query results quickly, and performance of additional adjustments to search parameters and drill-down through data. In another embodiment, using only conventional personal computer technology for analysis, queries can take only 30 seconds or less to process, with many requiring substantially less time. Data analyzer 308 provides near real-time or better in-depth reporting and search capabilities on the processed data set. Queries on individual locations, target audience versus individual locations, and searches based on target audience are all supported. As a result, sales staff of users 110 can plan targeted ad campaigns and report on previous campaign results in near real-time. Demographic and behavioral information is available on a per location basis. Demographic and behavioral information is also available on target audience subsets at a location.

The architecture is configured to interact with confidential data of users 110, e.g., ISPs, etc., and can ensure that the information remains confidential. In one embodiment of the invention, no confidential information is stored on a web server that users 110 interact with. In another embodiment, the web server can forward all user 110 requests to the architecture that can handle all interactions with the confidential information. The architecture can check and verify that users 110 are correctly logged in before handling a request. In addition, the architecture can work with a firewall of a user 110. The architecture can be behind the firewall where it can remain well protected. All interactions by user 110 with user interface 108 can be logged in and verified, in one embodiment of the invention. User interface 108 can run on a computer as described further below, with reference to FIG. 6, following the description of FIGs. 8-11, and 7.

An exemplary embodiment of the user interface 108 is described below with reference to FIGs. 7, and 8-11. FIG. 7 is discussed further below, following the description of FIG. 11. An exemplary reporting tool is now further described with reference to FIGs. 8-11.

FIG. 8 depicts block diagram 800 including an Internet browser 802, an activity monitor 804, a report server 806, and a report display 808. Internet browser 802 monitors client user activity such as, e.g., observing a location browsed by a client user, and the locations linked to by that location. A typical client user could include, e.g., a producer of content researching an

audience for the location browsed. Other client users can include, e.g., an advertising sales person looking for a specific target audience and an advertiser looking for a specific target audience. Activity monitor 804 can monitor the Internet browser 802, using, e.g., a separate browser window, a separate application or separate applet, a plug-in module installed into the browser, and a model incorporated into the browser.

Activity monitor 804 can then forward a query including, e.g., the location browsed and the locations linked to by the location browsed, to the report server 806. Report server 806 can then perform, e.g., processing functions, such as, e.g., generating a report for display by the report display 808. Report server 806 can provide a demographic and behavioral breakdown of an audience by the location. Report server 806 can also provide a targeted demographic and behavioral breakdown of an audience subset of the location. Report server 806 can also provide historical traffic levels for the location. Report server 806 can also provide a predicted future traffic availability for the location. Report server 806 can also provide audience analysis for the location and the locations to which it links.

Report server 806 can then send, e.g., a report based on the query, to the report display 808 for display. A report query can include the steps of sending the location from the internet browser 802, to activity monitor 804, and on to report server 806, sending a plurality of preferences of the requested information, generating a report on the report server 806 and receiving the report at the report display 808 for display, from the report server 806.

Report display 808 can display various statistical summary results of client user activity of the location browsed by the user of internet browser 802. Advantageously, report display 808 can provide detailed tracked activity statistics summarized to the level of the location being viewed using the internet browser 802. Report display 808 can display the results, e.g., using such tools as, e.g., a frame of the internet browser 802, a separate internet browser 802 window, and a separate applet. Advantageously, the summarized user behavior information can be obtained using the processes outlined in the second cross-referenced application. The demographic and behavioral analysis system architecture of the second cross-referenced application is reviewed below with reference to FIG. 1 of the present invention. The demographic and behavior analysis system of FIG. 1, above, provides analyzed information

tracking client usage on a per location basis, including, e.g., identifying, tracking and understanding user behavior on the Internet and in traditional stores.

FIG. 9 depicts an example report display 808 according to the present invention. Report display 808 can include, e.g., a control panel portion 902 and a report panel portion 904. Other panel portions can also be included in report display 808 such as, e.g., buttons, graphical charts, statistics including totals, subtotals, percentages, categories, demographics, target demographics, location identifiers, confidence ratings, filters, title bars, control, target, traffic and help icons.

FIG. 10 depicts an example embodiment of a demographic report 808a report display 808 according to the present invention. Demographic report 808a illustratively depicts control panel 902a and report panel 904a. Control panel 902a can enable selecting a location universal resource locator (URL) and controlling several example parameters associated with the report. Report panel 904a can display the statistical usage information for the location and parameters chosen using control panel 902a.

Control panel 902a can include in one embodiment, title 1002, demographic report, one or more buttons, such as, e.g., target button 1004 (for specifying a target audience), traffic report button 1006 (for viewing traffic statistics), and help button 1008.

Control panel 902a can also include a copy button 1010 which can enable a user to store a location's URL for later use.

Control panel 902a can include a demographic filter 1012 field that can narrow the range of demographic attributes to be displayed in the report panel 904a. For example, in one embodiment of the invention, in the range of 200-300, or more, demographics can be available for analysis for a given location. Suppose, for example, that of the 200-300 demographic attributes, only around 10 contained data of values greater than 5%, rather than listing all the demographic attributes, demographic filter 1012 can narrow the displayed list to, e.g., only values of greater than 5%. A simple data entry pull-down field permits the user to easily perform ad hoc trial and entry selections by merely selecting a value in filter field 1012 and then selecting the apply button 1014.

Control panel 902a can also include a location field, collectively illustrated as location fields 1016a and 1016b. Referring back to FIG. 8, in one embodiment, the location field 1016a and 1016b can automatically be filled in, according to the current location being viewed by the

user using input from the Internet browser 802 and activity monitor 804. In another embodiment of the invention, a location of interest to a user can be entered directly into, e.g., a location field 1016b, to view statistics on the location of interest. Further, in another embodiment, the location of interest entered into field 1016b can automatically cause internet browser 802 to open a browser window to view content at that location.

Control panel 902a can include confidence fields 1018a and 1018b which can provide information regarding a confidence level in the data provided in report panel 904a for the given location in field 1016b. In one embodiment of the invention, confidence data can be based on the size of audience having visited the location. The data can be scaled or normalized based on other similar representative sites, based on pure observed page hits, or based on other criteria. For example, if only 10 total persons have viewed a given location, this would be indicative of a lower level of confidence in the demographic data provided, as compared to a location where 1000 client users have viewed the site. In an alternative embodiment, if a search facility is included in control panel 902a, confidence field 1018b can be used to indicate the confidence in the search results.

Control panel 902a can also include target demographics fields 1020 and 1022. Target demographics field 1022 can display a list of targeted demographic attribute types, for which subtotal data can be provided. Report display 808a includes no selected target audience. If a user wanted to target a specific type of audience, the user could select one of the listed demographic attributes ("demographics") in report panel 904a, such as, e.g., 1032a, 1034a, 1036a, and so on, through 1052a. In one embodiment of the invention, which ever targeted demographics were selected could be displayed in field 1022. Selected targeted demographics can also be deselected, i.e. removed from the targeted demographics list, by selecting a targeted demographic in field 1022, in one embodiment.

In an embodiment of the invention, any and all fields can provide display functionality of report panel 904a and any and all fields can also be used to provide control functions of control panel 902a. For example, data associated with the targeted demographics selected, can be displayed in field 1022 and thus field 1022 can be thought of as part of report panel 904a, as well as part of control panel 902a.

In one embodiment of the invention, report panel 904a can include display of, e.g., other demographics 1024, which can in an embodiment of the invention, display demographics in column 1026, the percentage of client users tracked as visiting the location in columns 1028 and 1030. Column 1028 can display the percentage information, e.g., in the form of a histogram, a bar graph, a pie graph, and other graphical, numerical or other iconic representation of relative value. Column 1030, although illustrating a numerical representation of the value of the demographic percentages, can also illustrate the data in another form, such as, e.g., in the form of a histogram, a bar graph, a pie graph, and other numerical and other graphical or other iconic representation of relative value.

In one embodiment, demographics can be grouped according to related types of demographics, such as, e.g., age based, or gender based, demographics can be listed together, and sorted for ease of comparative review. Illustratively, gender demographics for male 1032a and female 1034a can be placed adjacent in order to permit improved readability and analysis, as shown of related percentage data 1032b, 1032c and 1034b, 1034c.

Similarly, age based demographics 1036a through 1042a can be placed adjacent one another in one embodiment, and can be sorted in numerical order.

Other demographics groupings can be organized adjacent to one another for ease of viewing. An example is the high level Internet domain of client users, such as, e.g., ".com," ".gov," ".net," ".org." Other large demographic populations such as, e.g., client users from Internet service providers or online service providers, such as, e.g., America Online, i.e. aol.com can also be listed as a separate category.

In one embodiment, gender based and age based demographics 1032a-1042a can be placed at the top of the other demographics 1024 list for ease of reading.

In one embodiment of the invention, the values of report panel 904a are automatically sorted before display by the value of column 1030. In another embodiment, the data is sorted including adjacent groupings such as gender and age based demographics groups 1032a-1042a, above.

In one embodiment, by selecting a column header 1026, 1028 or 1030, the data can be sorted by the selected column.

In another embodiment of the invention, the list of demographics are fixed and not necessarily in an alphabetical order.

In one embodiment, to target a specific demographic, a user can select one or more demographic categories and can then select the target audience button 1004. In another embodiment, the user can select a demographic group by another method of selection, such as, e.g., selecting a demographic and double clicking on it, or clicking with a right mouse button on a demographic and selecting target audience based on demographic, or selecting a demographic and dragging it to the target demographics fields 1020, 1022, or selecting several demographics and similarly selecting a targeted audience. Suppose, for example, that a user selects a demographic group including all user activity at a /rec/woodworking location 1016b, that is from Internet domain ".com" 1044a. The results of such a selection are illustrated below with reference to FIG. 11.

FIG. 11 depicts an example targeted demographic report 808b report display 808 according to the present invention. A user of the present invention could reach the screen as described, e.g., in the preceding paragraph. Targeted demographic report 808b illustratively depicts control panel 902b and report panel 904b. Control panel 902b can enable selecting a location universal resource locator (URL) and controlling several example parameters associated with the report. In one embodiment of the invention, control panel 902b includes only the title bar area including title 1102 and buttons 1104, 1106 and 1108. In another embodiment of the invention, control panel 902b can include any area of report 808b which can be used to control the data output in report panel 904b. Report panel 904b can display the targeted demographic statistical usage information for the location and parameters chosen using control panel 902b. In one embodiment, report panel 904b can include portions of report 808b which are also included as portions of 902b.

Control panel 902b can include in one embodiment, title 1102, demographic report, one or more buttons, such as, e.g., target button 1104 (for specifying a target audience, used to reach targeted demographic page 808b), traffic report button 1106 (for viewing traffic statistics), and help button 1108.

Control panel 902b can also include a copy button 1110 which can enable a user to store a location's URL for later use.

Control panel 902b can include a demographic filter 1112 field which can narrow the range of demographic attributes to be displayed in the report panel 904b. For example, in one embodiment of the invention, in the range of 200-300, or more, demographics can be available for analysis for a given location. Suppose, for example, that of the 200-300 demographic attributes, only around 10 contained data of values greater than 5%, rather than listing all the demographic attributes, demographic filter 1112 can narrow the displayed list to, e.g., only values of greater than 5%. A simple data entry pull-down field permits the user to easily perform ad hoc trial and entry selections by merely selecting a value in filter field 1112 and then selecting the apply button 1114.

Control panel 902b can also include a location field, collectively illustrated as location fields 1116a and 1116b. Referring back to FIG. 8, in one embodiment, the location field 1116a and 1116b can automatically be filled in, according to the current location being viewed by the user using input from the Internet browser 802 and activity monitor 804. In another embodiment of the invention, a location of interest to a user can be entered directly into, e.g., a location field 1116b, to view statistics on the location of interest. Further, in another embodiment, the location of interest entered into field 1116b can automatically cause internet browser 802 to open a browser window to view content at that location.

Control panel 902b can include confidence fields 1118a and 1118b which can provide information regarding a confidence level in the data provided in report panel 904b for the given location in field 1116b. In one embodiment of the invention, confidence data can be based on the size of audience having visited the location. The data can be scaled or normalized based on other similar representative sites, based on pure observed page hits, or based on other criteria. For example, if only 10 total persons have viewed a given location, this would be indicative of a lower level of confidence in the demographic data provided, as compared to a location where 1000 client users have viewed the site. In an alternative embodiment, if a search facility is included in control panel 902b, confidence field 1118b can be used to indicate the confidence in the search results.

Control panel 902b can also include target demographics fields 1120 and 1122a through 1122d. Target demographics field 1122a-1122d can provide similar column headings for targeted demographics 1144a, 1144b, 1144c and 1144d for targeted demographic ".com" and can

36731/161794 (A-21415.PCT)

display data for the targeted demographic attribute including subtotaled data and graphical or numerical information about the targeted demographic. Demographic data field 1144c can indicate the percentage of total users for the location which fall within the targeted demographic group. Target demographic data field 1144d can include the percentage of total users for the location which fall also fall within the targeted demographic group 1144a which is, in this case, the same as the data in field 1144c.

Report display 808b includes the "com" target audience. If a user wanted to target a specific type of audience, the user could select one of the other listed demographic attributes ("demographics") in report panel 904b, such as, e.g., 1132a, 1134a, 1136a, and so on, through 1152a, in addition to targeted demographic 1144a.

In one embodiment of the invention, targeted demographics previously selected can be displayed below field 1122a-1122d. Selected targeted demographics can also be deselected, i.e. removed from the targeted demographics list, by deselecting a selected targeted demographic in field 1122, in one embodiment.

In an embodiment of the invention, any and all fields of report 808b can provide display functionality of report panel 904b and any and all fields of report 808b can also be used to provide control functions of control panel 902b. For example, data associated with the targeted demographics selected, can be displayed below field 1122 and thus field 1122 can be thought of as part of report panel 904b, as well as part of control panel 902b.

In one embodiment of the invention, report panel 904b can include display of, e.g., other demographics 1124, which can in an embodiment of the invention, display, e.g., demographics in column 1126, the percentage of client users tracked as visiting the location (including an additional indication of those users who also are members of the targeted demographic or demographics) in columns 1128, 1130 and 1131. Column 1128 can display the percentage information indicating the portion of users in the demographic only, and the portion of users in both the demographic and the targeted demographics. The data can be provided in two separate forms (not shown) or integrated (as shown) using different colors, e.g., in the form of a histogram, a bar graph, a pie graph, and other graphical, numerical or other iconic representation of relative value. Column 1130, although illustrating a numerical representation of the value of the demographic percentages, can also illustrate the data in another form, such as, e.g., in the

form of a histogram, a bar graph, a pie graph, and other numerical and other graphical or other iconic representation of relative value. Column 1131 can include similar data/information showing the percentage of users of the location 1116b which are members of demographic 1126 and targeted demographic 1144a. For example, for a male demographic type 1132a, a percentage of total client users of location 1116b is shown numerically in field 1132c and are graphed as part (the longer histogram) of field 1132b. Similarly, for male demographic type 1132a which are also members of targeted demographic type 1144a, a percentage of the total users of location 1116d meeting both targeted and group demographics is listed in field 1132c and is graphed as the shorter graph in field 1132b. In one embodiment, the graphical representations of column 1128 can include multiple colors, such as, e.g., blue for the shorter and yellow for the longer bar in field 1132b.

In one embodiment, demographics can be grouped according to related types of demographics, such as, e.g., age based, or gender based, demographics can be listed together, and sorted for ease of comparative review. Illustratively, gender demographics for male 1132a and female 1134a can be placed adjacent in order to permit improved readability and analysis, as shown of related percentage data 1132b, 1132c, 1132d and 1134b, 1134c, and 1134d.

Similarly, age based demographics 1136a through 1142a can be placed adjacent one another in one embodiment, and can be sorted in numerical order.

Other demographics groupings can be organized adjacent to one another for ease of viewing. An example is the high level Internet domain of client users, such as, e.g., ".com," ".gov," ".net," ".org." Other large demographic populations such as, e.g., client users from Internet service providers or online service providers, such as, e.g., America Online, i.e. aol.com can also be listed as a separate category in field 1148a, for example. Where individual ".com" or ".net" types are listed, other ".com" 1150b and other ".net" 1152b demographics can also be provided.

In one embodiment, gender based and age based demographics 1132a-1142a can be placed at the top of the other demographics 1124 list for ease of reading.

In one embodiment of the invention, the values of report panel 904b are automatically sorted before display by the value of column 1130. In another embodiment, the data is sorted

including adjacent groupings such as gender and age based demographics groups 1132a-1142a, above.

In one embodiment, by selecting a column header 1126, 1128, 1131 or 1130, the data can be sorted by the selected column.

In another embodiment of the invention, the list of demographics are fixed and not necessarily in an alphabetical order.

In one embodiment, to target a specific demographic, a user can select one or more demographic categories and can then select the target audience button 1104. In another embodiment, the user can select a demographic group by another method of selection, such as, e.g., selecting a demographic and double clicking on it, or clicking with a right mouse button on a demographic and selecting target audience based on demographic, or selecting a demographic and dragging it to the target demographics fields 1120, 1122 and 1124, or selecting several demographics and similarly selecting a targeted audience.

A user can select multiple demographics for target by clicking on the demographics in column 1126 to select them. If multiple members of a mutually exclusive group are selected, they are logically or'ed together and and'ed with the remaining target demographics.

A user can also select to display a traffic report by selecting button 1106. It should be apparent to those skilled in the art, that the use of the expression "select" as used in this application can include the use of, e.g., a mouse pointer, button, touchpad, pointing device, touchscreen, key, cursor or other known selection device. Suppose, for example, that a user selects to display traffic report statistics using button 1106. The results of such a selection are illustrated below with reference to FIG. 7.

FIG. 7 depicts an exemplary traffic report 808c report display 808 illustrating an example range of traffic summarized on an example weekly basis in an embodiment of the present invention. A user of the present invention could reach the screen as described, e.g., in the preceding paragraph. Traffic report 808c illustratively depicts control panel 902c and report panel 904c. Control panel 902c can enable selecting a location universal resource locator (URL) and controlling several example parameters associated with the report. In one embodiment of the invention, control panel 902c includes only the title bar area including title 702 and buttons 704, 706 and 708. In another embodiment of the invention, control panel 902c can include any

area of report 808c which can be used to control the data output in report panel 904c. Report panel 904c can display the targeted demographic statistical usage information for the location and parameters chosen using control panel 902c. In one embodiment, report panel 904c can include portions of report 808c which are also included as portions of 902c.

Control panel 902c can include in one embodiment, title 702, traffic report, one or more buttons, such as, e.g., demographics button 704 (for viewing a demographic report 808a) and help button 708.

Control panel 902c can include time span field 1112 which can narrow the range of traffic to be displayed in the report panel 904c. For example, in one embodiment of the invention, traffic can be totaled, e.g., in daily, weekly, monthly, yearly increments. A simple data entry pull-down field permits the user to selecting a time span in field 1112 and then select the apply button 714. A starting date 706 and ending date 710 for the range can also be selected. Calendar buttons 705 and 709 permit graphical selection of start and end dates, in one embodiment.

Control panel 902c can also include a location field, collectively illustrated as location fields 716a and 716b. Referring back to FIG. 8, in one embodiment, the location field 716a and 716b can automatically be filled in, according to the current location being viewed by the user using input from the Internet browser 802 and activity monitor 804. In another embodiment of the invention, a location of interest to a user can be entered directly into, e.g., a location field 716b, to view statistics on the location of interest. Further, in another embodiment, the location of interest entered into field 716b can automatically cause internet browser 802 to open a browser window to view content at that location.

Control panel 902c can also include traffic fields 720 and 722a through 722d. Traffic fields include an include field 718, a date field 720, a traffic field 722, total fields 724 and 726. Each time span can then be provided, in the illustrated example, each week, shown in fields 728b through 736b, and can be selected for inclusion in a separate total field such as, for example, total 738 and 740, using selection fields 728a-736a. Weekly data can appear in fields 728c through 736c. Partial time windows can be indicated in one embodiment as shown with text 742. Timestamp information can be included, such as, e.g., report time 744.

In one embodiment the date range is inclusive, based on US date systems, other date systems can be used.

In one embodiment, if more data appears than can fit comfortably on a page, then scroll bars can appear.

In another embodiment, a total field can be listed at the top and bottom of a long list as shown.

In another embodiment, audience analysis can have its own control panel 902 and report panel 904. The audience analysis control panel can include a method for selecting an analysis type, such as, e.g., a correlation or a bayesian. The audience analysis report panel can include, e.g., the locations that score highest (or lowest) using the currently selected analysis method. The locations can be, e.g., the list of locations linked to by the browsed location, or the list of all locations.

FIG. 6 depicts an exemplary computer system. Specifically, FIG. 6 illustrates an example computer 600 in a preferred embodiment is a personal computer (PC) system running an operating system such as Windows 98, OS/2, Mac/OS, or UNIX. However, the invention is not limited to these platforms. Instead, the invention can be implemented on any appropriate computer system running any appropriate operating system, such as Solaris, Irix, Linux, HP/UX, OSF, Windows 98, Windows NT, OS/2, Mac/OS, and any others that can support Internet access.

In one embodiment, the present invention is implemented on a computer system operating as discussed herein. An exemplary computer system, computer 600 is shown in FIG. 6. Other components of the invention, such as client workstations, proxy servers, network communication servers, remote access devices, client computers, server computers, routers, web servers, data, media, audio, video, telephony or streaming technology servers could also be implemented using a computer such as that shown in FIG. 6.

The computer system 600 includes one or more processors, such as processor 602. The processor 602 is connected to a communication bus 604.

The computer system 600 also includes a main memory 606, preferably random access memory (RAM), and a secondary memory 608. The secondary memory 608 includes, e.g., a hard disk drive 610 and/or a removable storage drive 612, representing a floppy diskette drive, a

magnetic tape drive, a compact disk drive, etc. The removable storage drive 612 reads from and/or writes to a removable storage unit 614 in a well known manner.

Removable storage unit 614, also called a program storage device or a computer program product, represents a floppy disk, magnetic tape, compact disk, etc. The removable storage unit 614 includes a computer usable storage medium having stored therein computer software and/or data, such as an object's methods and data.

Computer 600 also includes an input device such as (but not limited to) a mouse 616 or other pointing device such as a digitizer, and a keyboard 618 or other data entry device.

Computer 600 can also include output devices, such as, e.g., display 620. Computer 600 can include input/output (I/O) devices such as, e.g., network interface cards 622 and modems 150 and 152.

Computer programs (also called computer control logic), including object oriented computer programs, are stored in main memory 606 and/or the secondary memory 608 and/or removable storage units 614, also called computer program products. Such computer programs, when executed, enable the computer system 600 to perform the features of the present invention as discussed herein. In particular, the computer programs, when executed, enable the processor 602 to perform the features of the present invention. Accordingly, such computer programs represent controllers of the computer system 600.

In another embodiment, the invention is directed to a computer program product comprising a computer readable medium having control logic (computer software) stored therein. The control logic, when executed by the processor 602, causes the processor 602 to perform the functions of the invention as described herein.

In yet another embodiment, the invention is implemented primarily in hardware using, e.g., one or more state machines. Implementation of these state machines so as to perform the functions described herein will be apparent to persons skilled in the relevant arts.

While various embodiments of the present invention have been described above, it should be understood that they have been presented by way of example only, and not limitation. Thus, the breadth and scope of the present invention should not be limited by any of the above-described exemplary embodiments, but should be defined only in accordance with the following claims and their equivalents.

What is Claimed is:

1. A method for analyzing user activity using an inventory-centric approach, comprising the steps of:

- (1) accessing raw user data;
- (2) processing said raw user data to generate clean user data; and
- (3) processing said clean user data using a core technology to generate inventory-centric aggregated user data.

2. The method according to claim 1, wherein said step (1) comprises identifying and tracking a user accessing the Internet via a proxy server, comprising the steps of:

- (a) accessing a proxy log;
- (b) accessing an IP address assignment log; and
- (c) merging said proxy log and said IP address assignment log to obtain virtual cookie identification data.

3. The method according to claim 2, wherein said proxy log comprises a proxy log data record including the following fields:

a location requested by the user, a first IP address of the user making the request, an action requested by the user, and a time of the request.

4. The method according to claim 2, wherein said IP address assignment log comprises an IP address assignment log data record including the following fields:

a second IP address assigned to the user, a userID of the user, and a time window of assignment of said second IP address to the user.

5. The method according to claim 3, wherein said IP address assignment log comprises an IP address assignment log data record including the following fields:

a second IP address assigned to the user, a userID of the user, and a time window of assignment of said second IP address to the user.

6. The method according to claim 2, wherein said virtual cookie identification data comprises:

a location, an action, and a userID.

7. The method according to claim 5, wherein said virtual cookie identification data comprises:

said location, said action, and said userID.

8. The method according to claim 7, wherein said step (c) includes:

correlating said first IP address and said second IP address, and said time of the request and said timewindow of the assignment to determine said userID making the request.

9. The method according to claim 2, further comprising the step of:
outputting said virtual cookie identification data.

10. The method according to claim 2, wherein said method is performed at least one of post-browsing and real-time.

11. The method according to claim 2, further comprising the step of:

(e) analyzing said virtual cookie identification data.

12. The method according to claim 11, wherein said step (e) comprises at least one step of the following steps of:

(i) analyzing demographic data using said virtual cookie identification data;

and

(ii) analyzing psychographic data using said virtual cookie identification data.

13. The method according to claim 12, wherein step (i) includes associating said demographic data with said userID.
14. The method according to claim 12, wherein said step (ii) includes associating said psychographic data with said userID.
15. The method according to claim 2, wherein said proxy server is at least one of owned, leased and operated by an Internet service provider (ISP).
16. The method according to claim 2, wherein said proxy server is at least one of owned, leased and operated by a corporate network.
17. The method according to claim 2, wherein said proxy server is at least one of a caching technology and a logging technology that can observe and record activity of the user, and wherein said proxy log is a log of at least one of said caching technology and said logging technology.
18. The method according to claim 2, wherein said IP address assignment log is at least one of:
 - a dial-up log;
 - a dynamically assigned IP address log;
 - a dynamic host configuration protocol (DHCP) compliant remote access device (RAD) log; and a statically assigned IP address log.
19. The method according to claim 1, further comprising the step of:
 - (4) analyzing said inventory-centric aggregated user data.
20. The method according to claim 19 wherein said step (4) comprises displaying location-specific reports as a user browses the Internet, comprising the steps of:

- (a) browsing the Internet using a browser;
 - (b) monitoring activity with said browser;
 - (c) observing a location browsed wherein said location includes content;
 - (d) requesting a report on said location; and
 - (e) displaying said report regarding said location.
21. The method of claim 20, wherein said content is from a website.
22. The method of claim 20, wherein said content is at least one of static and being dynamically generated.
23. The method of claim 20, wherein said step (b) comprises:
- (i) monitoring using an activity monitor.
24. The method of claim 20, wherein said step (d) comprises:
- (i) requesting said report from a report server.
25. The method of claim 20, wherein said step (e) comprises:
- (i) displaying said report on a report display.
26. The method of claim 20, wherein said browser is an Internet browser application program.
27. The method of claim 20, wherein said browsing step is performed by a user.
28. The method of claim 27, wherein said user is at least one of the following:
- a producer researching audience for said location;
 - an advertising sales person looking for a specific target audience; and
 - an advertiser looking for a specific target audience.
29. The method of claim 23, wherein said step (i) comprises at least one of the following

steps of:

- (A) monitoring said browser using a separate browser window;
- (B) monitoring said browser using at least one of a separate application and a separate applet;
- (C) monitoring said browser using a plug-in module installed into said browser; and
- (D) monitoring said browser with a module incorporated into said browser.

30. The method of claim 20, wherein step (d) includes at least one of the following:

- (i) requesting a demographic and behavioral breakdown of an audience of said location;
- (ii) requesting a targeted demographic and behavioral breakdown of an audience subset of said location;
- (iii) requesting historical traffic levels for said location;
- (iv) requesting predicted future traffic availability for said location; and
- (v) requesting audience analysis for said location.

31. The method of claim 24, wherein said report server is running on at least one of the following:

- a computer of a user;
- a separate computer from said computer of said user; and
- said computer of said user having an activity monitor integrated with said report server.

32. The method of claim 20 wherein step (d) includes at least one step of the following steps of:

- (i) sending said location;
- (ii) sending a plurality of preferences of a user;
- (iii) generating said report on a report server; and
- (iv) receiving said report from said report server.

33. The method of claim 32, wherein said plurality of preferences of said user include at least one of the following:

- (A) a type of said report to be generated by said report server; and
- (B) a display preference determining how said report is to be displayed.

34. The method according to claim 1, wherein said raw user data includes at least one of the following:

user action records including at least one of a userID of a user, an action performed by said user, and a location where said user performed said action;

user demographics records including at least one of a userID of a user, and at least one demographic associated with said user; and

user records including at least one of a userID of a user and a name of said user.

35. The method according to claim 1, wherein step (3) comprises at least one step of the following steps:

(a) receiving said clean user data;

(b) accessing user action records, wherein each of said user action records includes at least one of a userID of a user, an action performed by said user, and a location where said user performed said action;

(c) identifying a plurality of said users for each of said locations wherein said plurality of users performed actions at each of said locations in said user action records; and

(d) generating said inventory-centric aggregated user data using said plurality of said users associated with said each of said locations.

36. The method according to claim 35, wherein step (d) includes at least one step of the following steps:

(i) receiving said plurality of said users and said clean user data associated with said plurality of said users;

(ii) generating a cluster membership for each of said users of said plurality of said users; and

(iii) aggregating said clean user data of each of said users by said cluster membership into said inventory-centric aggregated user data.

37. The method according to claim 36, wherein step (ii) includes at least one of the following steps:

(A) classifying said user by matching said clean user data of said user against definition; and

(B) clustering said user including grouping said user with substantially similar users based on similarities of said clean user data.

38. A method for enhancing analysis of user activity using aggregated user data by identifying and removing atypical users, comprising the steps of:

(1) accessing raw user data;

(2) detecting by their actions atypical users and removing said atypical users to generate clean user data; and

(3) processing said clean user data using a core technology to generate aggregated user data.

39. The method according to claim 38, further comprising the step of:

(4) analyzing said aggregated user data.

40. The method according to claim 38, wherein said raw user data includes at least one of the following:

user action records including at least one of a userID of a user, an action performed by said user, and a location where said user performed said action;

user demographics records including at least one of a userID of a user, and at least one demographic associated with said user; and

user records including at least one of a userID of a user and a name of said user.

41. The method according to claim 38, wherein step (2) comprises at least one of the following steps:

- (a) accessing user action records from said raw user data;
- (b) identifying said atypical users by scanning said user action records;
- (c) accessing said raw user data; and
- (d) removing said atypical users from said raw user data to generate said clean user data.

42. The method according to claim 41, wherein said atypical users includes at least one of the following:

- software robots; and
- staff personnel.

43. The method according to claim 41, wherein said scanning step includes the step of using statistical methods to identify said atypical users.

44. A method for enhancing analysis of user activity using aggregated user data by merging a plurality of consistent aggregated user data, comprising the steps of:

- (1) accessing raw user data;
- (2) processing said raw user data to generate clean user data;
- (3) processing said clean user data using a core technology to generate consistent aggregated user data; and
- (4) merging a plurality of said consistent aggregated user data into merged aggregated user data.

45. The method according to 44, further comprising the step of:

- (5) analyzing said merged aggregated user data.

46. The method according to claim 44, wherein said raw user data includes at least one of the following:

user action records including at least one of a userID of a user, an action performed by said user, and a location where said user performed said action;

user demographics records including at least one of a userID of a user, and at least one demographic associated with said user; and

user records including at least one of a userID of a user and a name of said user.

47. The method according to claim 44, wherein said plurality of said consistent aggregated user data is generated from a plurality of said raw user data from at least one of the following: different time periods;

different servers including at least one of web servers, ad servers, logging servers, and point of sale servers;

a subset of a larger set of said raw user data; and

raw user data of a previous year for seasonality.

48. The method according to claim 44, wherein step (3) comprises at least one step of the following steps:

(a) receiving said clean user data;

(b) generating a consistent cluster membership for each user in said clean user data; and

(c) aggregating said clean user data by said consistent cluster membership into said consistent aggregated user data.

49. The method according to claim 48, wherein step (c) includes at least one of the following steps:

(i) classifying said user by matching said clean user data of said user against a definition, wherein said definition remains constant during generation of said plurality of said consistent aggregated user data; and

(ii) clustering said user including grouping said user with substantially similar users based on similarities of said clean user data, wherein said groupings remain constant during generation of said plurality of said consistent

aggregated user data.

50. The method according to claim 44, wherein step (4) comprises at least one step of the following steps:

- (a) accessing a plurality of said consistent aggregated user data;
- (b) accessing auxiliary data; and
- (c) merging said plurality of said consistent aggregated user data and said auxiliary data to obtain said merged aggregated user data.

51. The method according to claim 44, wherein said auxiliary data includes:

date information recording types of said clean user data contained in each of said plurality of said consistent aggregated user data.

52. The method according to claim 51, wherein said types of said clean user data includes demographics.

53. The method according to claim 50, wherein step (c) includes at least one of the following steps:

- (i) averaging of said plurality of said consistent aggregated user data; and
- (ii) weighted averaging of said plurality of said consistent aggregated user data.

54. A method for analyzing user activity across the Internet for determining user behavior, comprising the steps of:

- (1) accessing raw user data for user activity on a user-defined plurality of sites visited by users of the Internet;
- (2) processing said raw user data to generate clean user data; and
- (3) processing said clean user data using a core technology to generate aggregated user data.

55. The method according to claim 54, further comprising the step of:

(4)analyzing said aggregated user data.

56. The method according to claim 54, wherein said raw user data includes the following:
user action records including at least one of a userID of a user, an action performed by said user, and a location where said user performed said action.

57. The method according to claim 54, wherein said raw user data is obtained using one of the following methods:

(a) observing said user activity in or near a network communication server wherein said network communication server can enable users to connect to the Internet; and

(b) observing said user activity using software on computers of said users.

58. The method according to claim 54, wherein step (2) comprises at least one step of the following steps:

(a) receiving said raw user data;

(b) accessing user action records including at least one of a userID of a user, an action performed by said user, and a location where said user performed said action;

(c) identifying said user action records substantially similar to behavioral demographic requirements; and

(d) generating said clean user data by associating said users from said user actions records with behavioral demographics associated with said behavioral demographic requirements.

59. The method according to claim 58, wherein said behavioral demographic requirements include at least one of the following:

an action indicating membership in a behavioral demographic;

a location indicating membership in a behavioral demographic; and

a location and action pair indicating membership in a behavioral demographic.

60. The method according to claim 54, wherein step (3) comprises at least one step of the following steps:

- (a) receiving said clean user data;
- (b) accessing user action records including at least one of a userID of a user, an action performed by said user, and a location where said user performed said action;
- (c) identifying a plurality of said users for each said location wherein said users performed actions at said location in said user action records; and
- (d) generating said aggregated user data using said plurality of said users associated with each said location.

61. The method according to claim 60, wherein step (c) includes at least one step of the following steps:

- (i) receiving said plurality of said users and said clean user data associated with said plurality of said users;
- (ii) generating a cluster membership for each said user in said plurality of said users; and
- (iii) aggregating said clean user data of users by said cluster membership into said aggregated user data.

62. The method according to claim 61, wherein step (ii) includes at least one of the following steps:

- (A) classifying said user by matching said clean user data of said user against a definition; and
- (B) clustering said user including grouping said user with substantially similar users based on similarities of said clean user data.

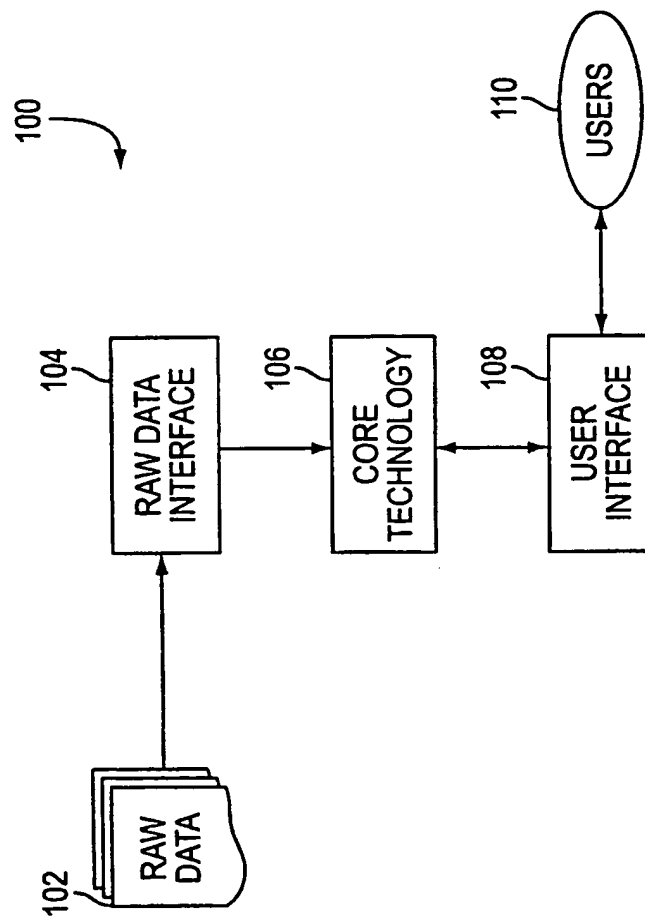


FIG. 1A

2/20

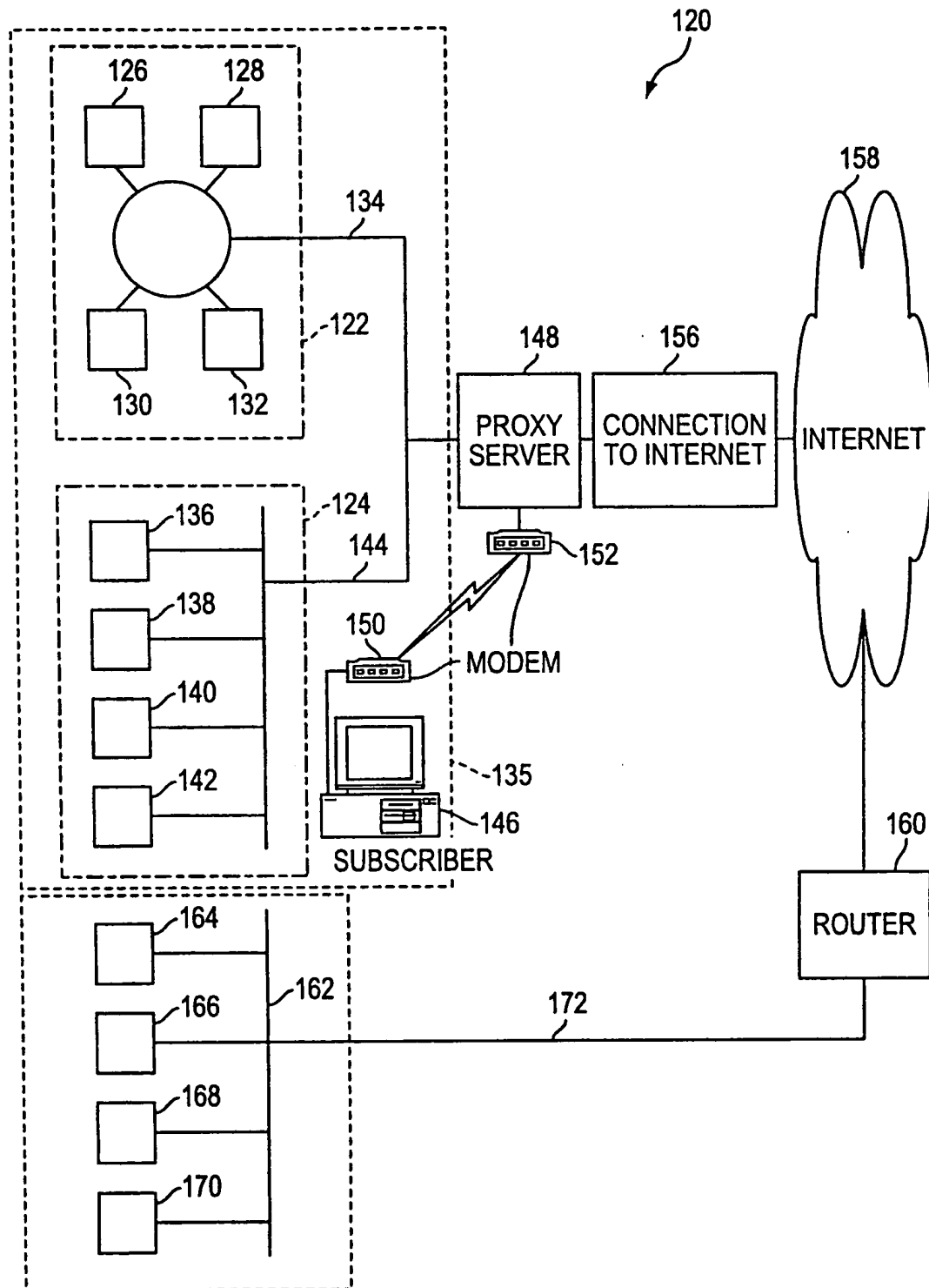


FIG. 1B

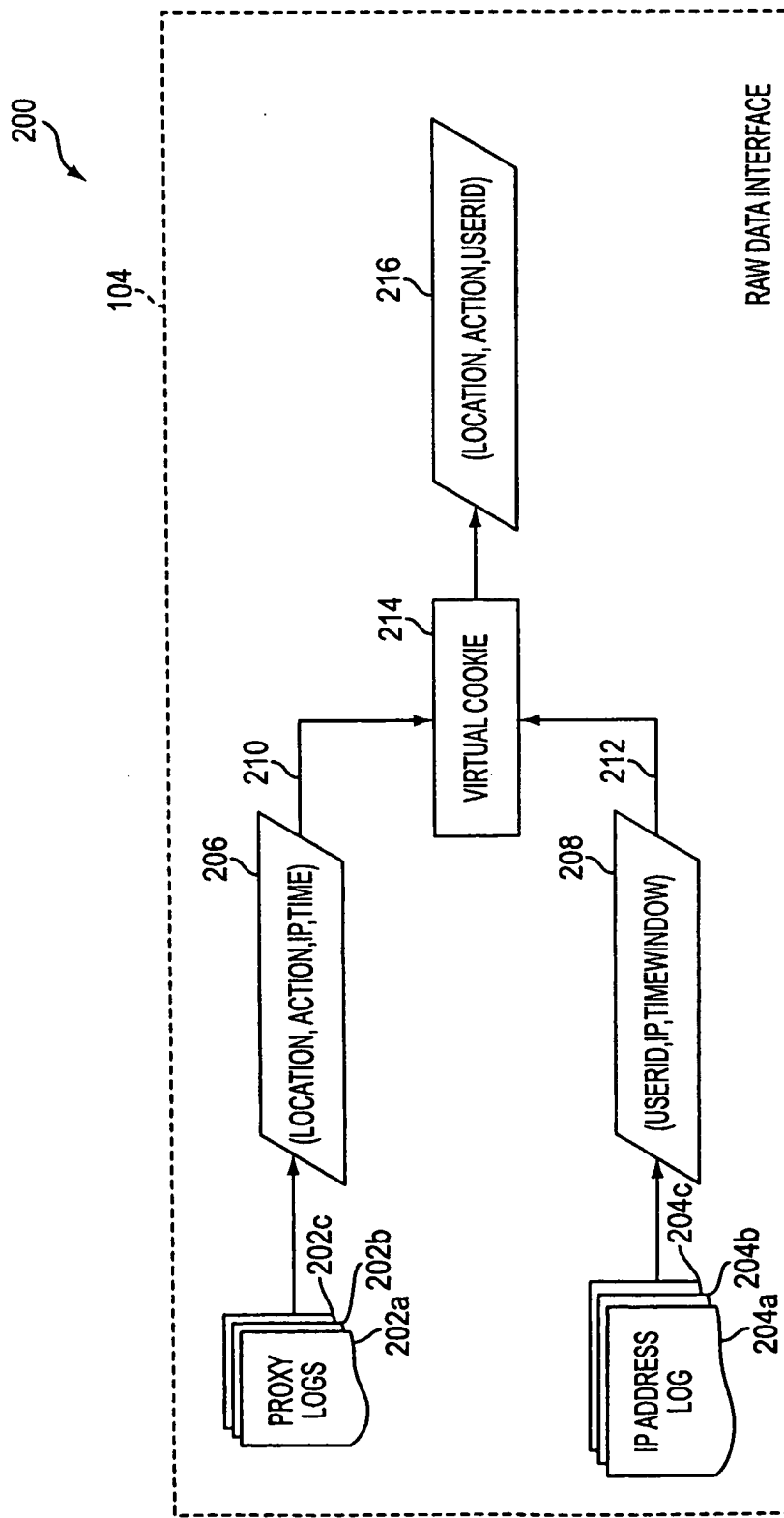


FIG. 2A

4/20

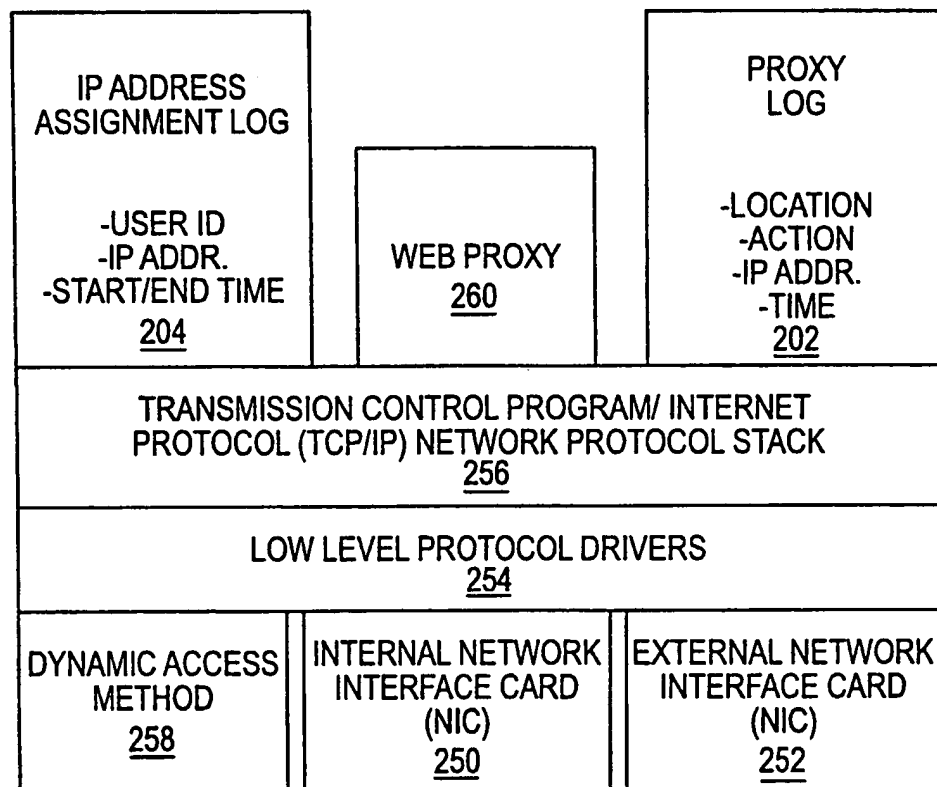


FIG. 2B

5/20

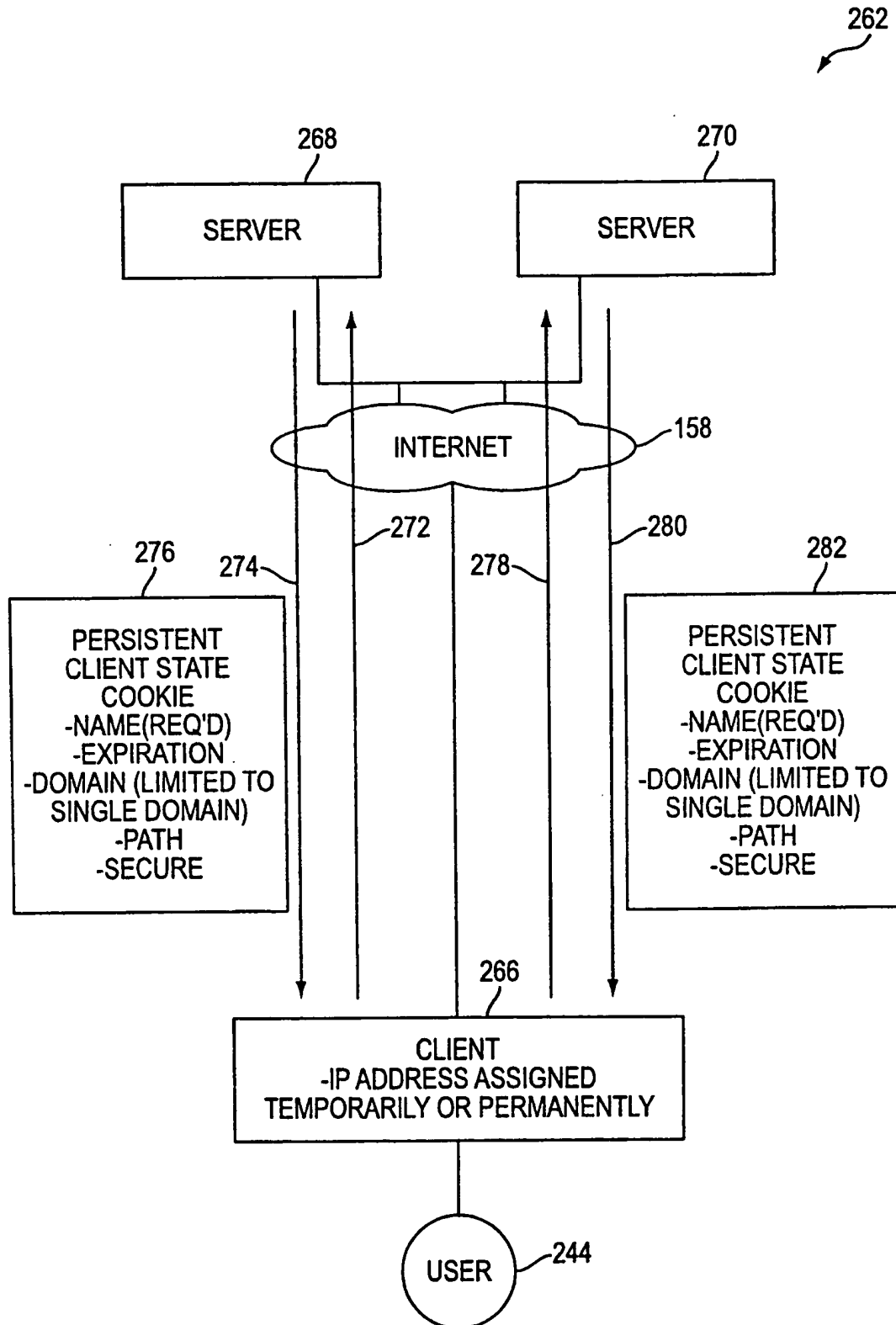


FIG. 2C

SUBSTITUTE SHEET RULE 92 BIS

6/20

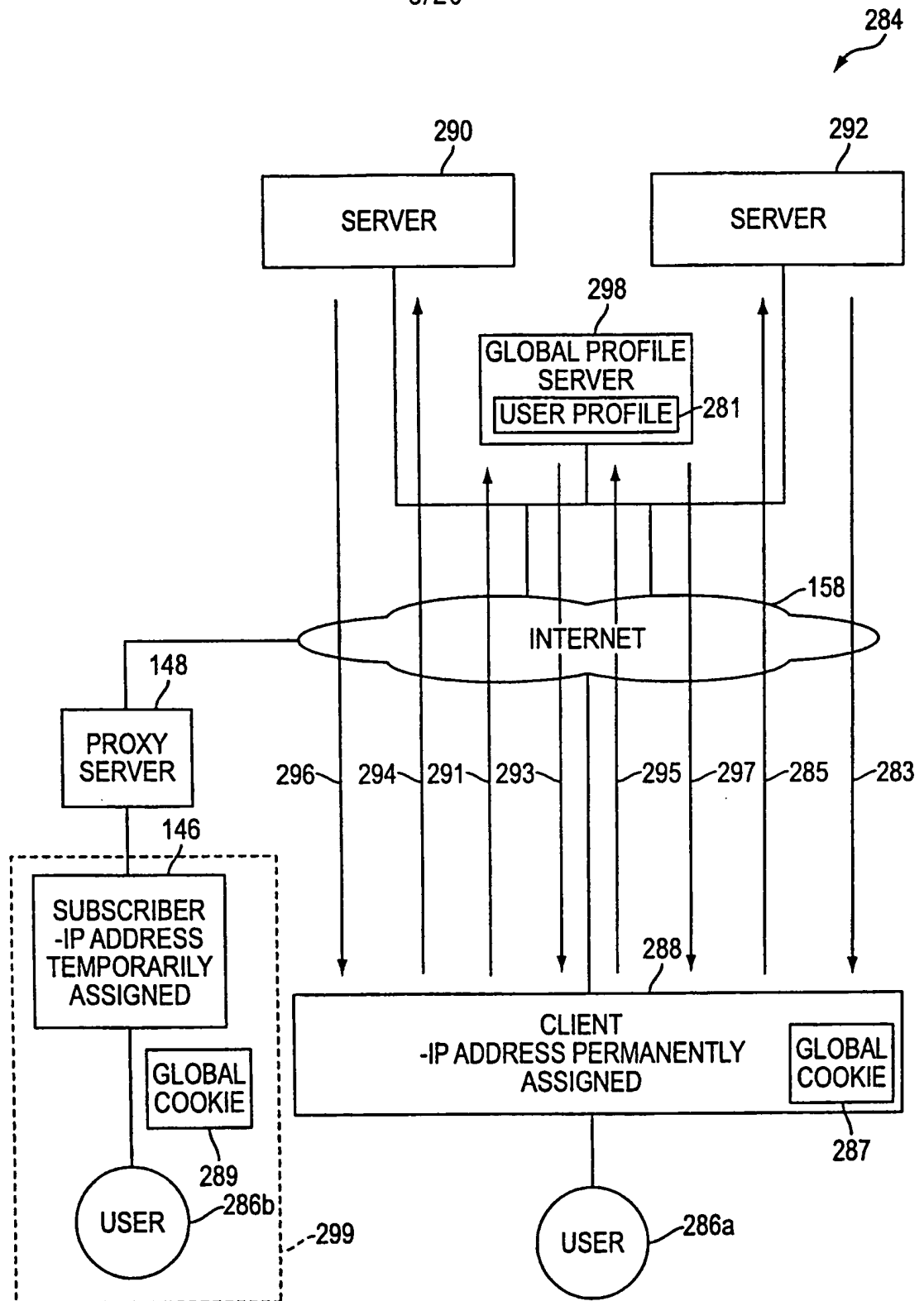


FIG. 2D

7/20

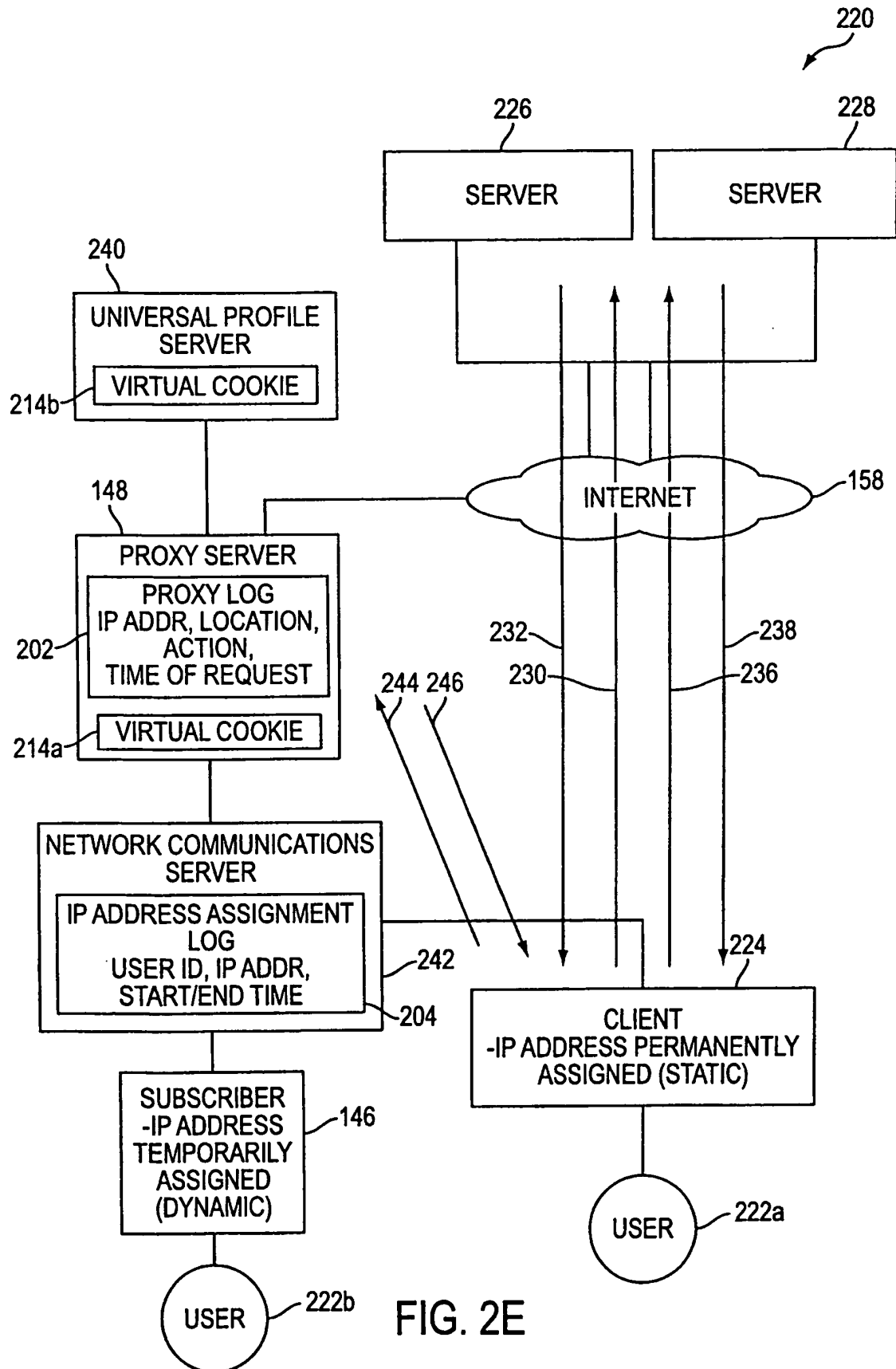


FIG. 2E

SUBSTITUTE SHEET RULE 92 BIS

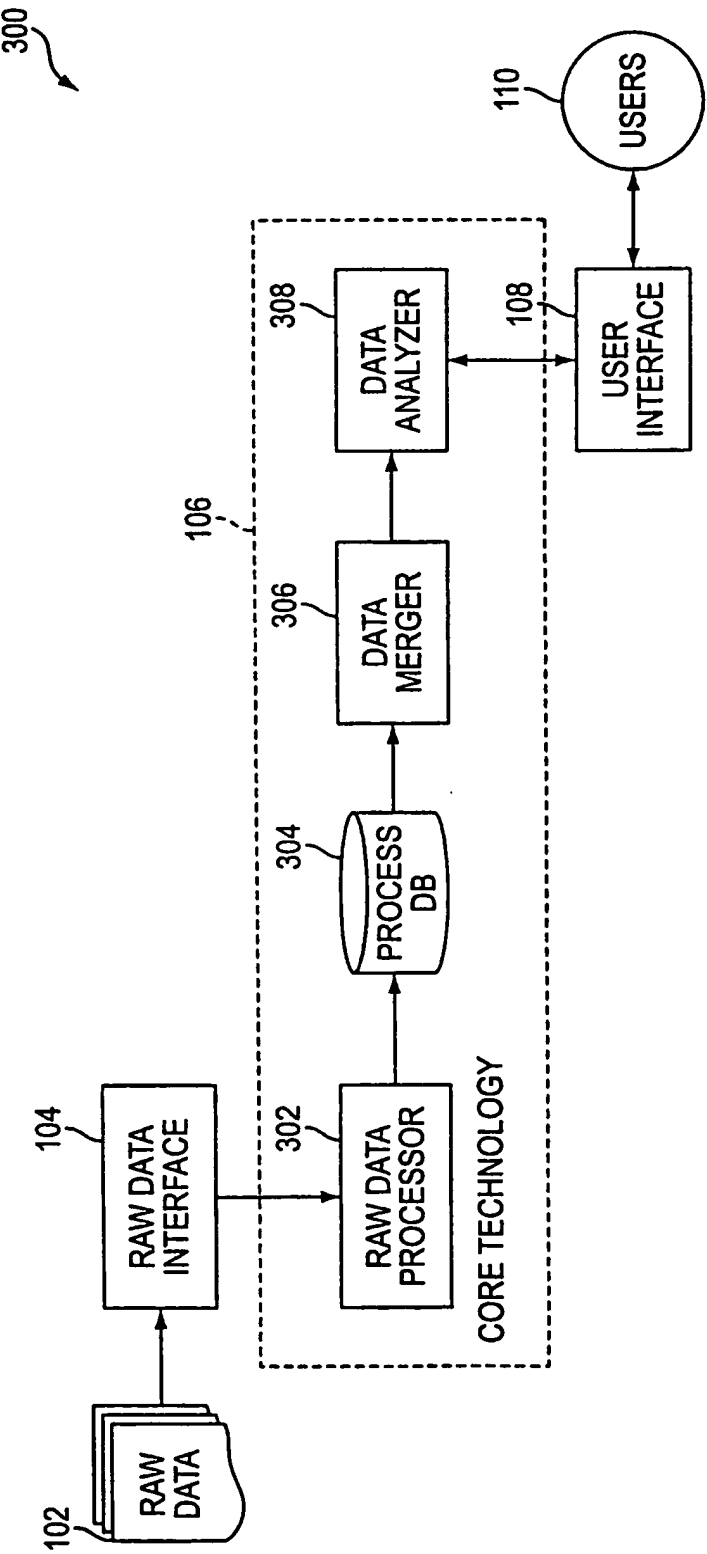


FIG. 3

304A

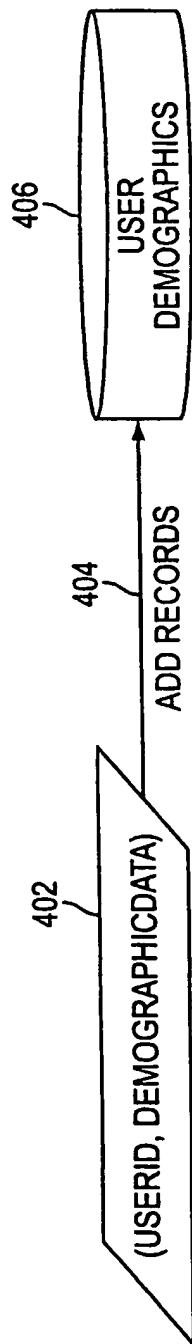


FIG. 4A

304B

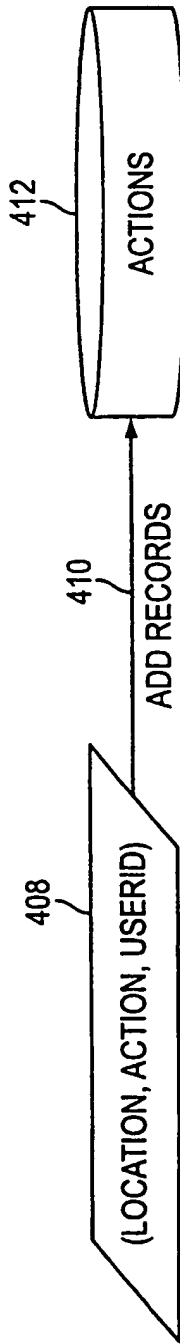


FIG. 4B

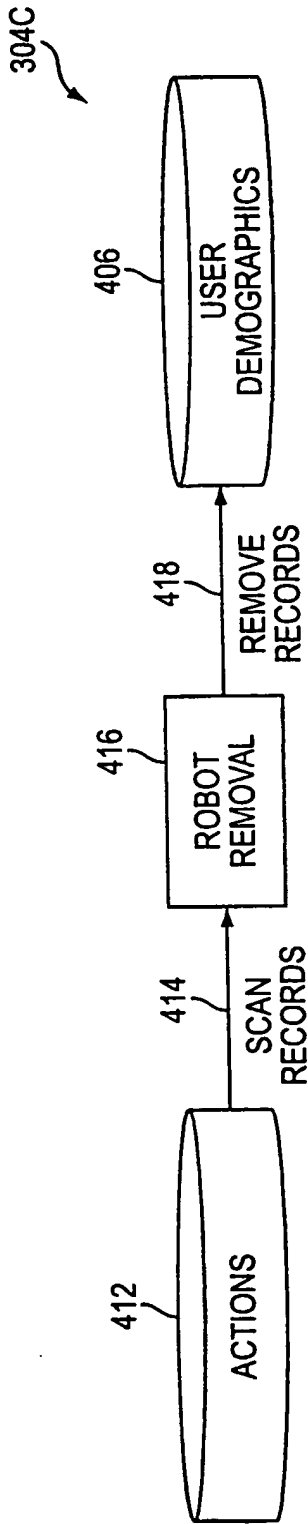


FIG. 4C

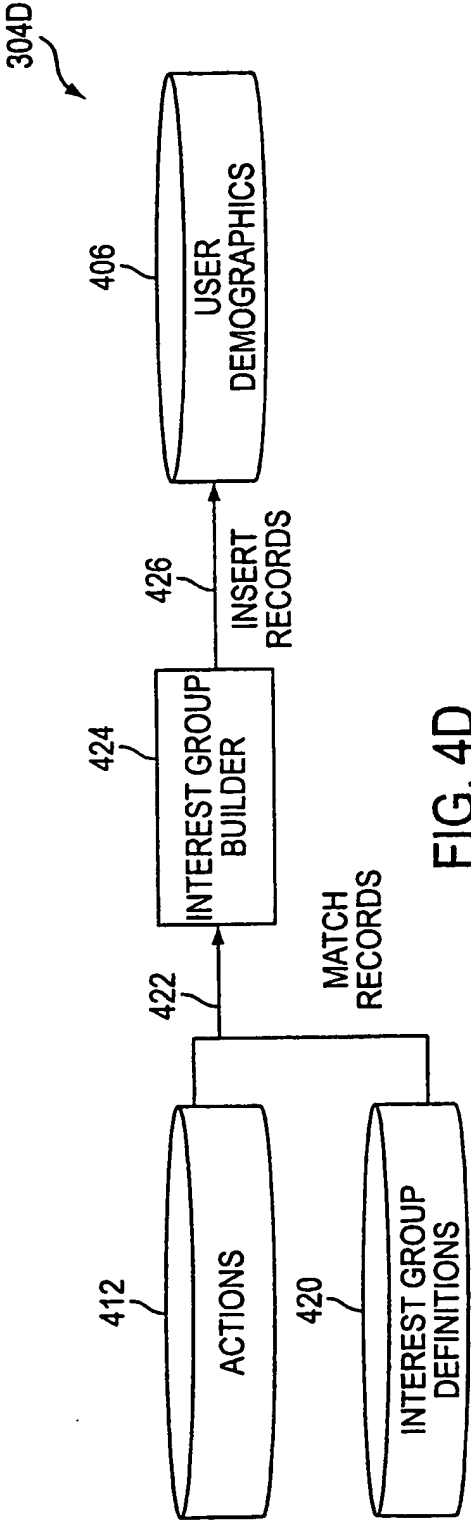


FIG. 4D

11/20

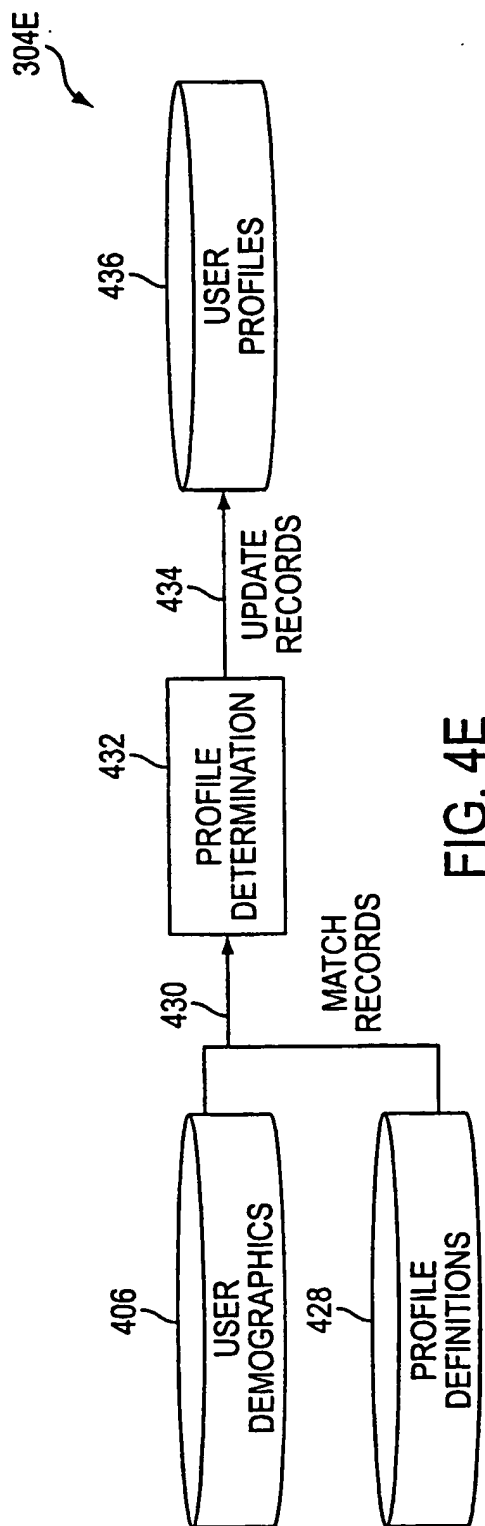


FIG. 4E

12/20

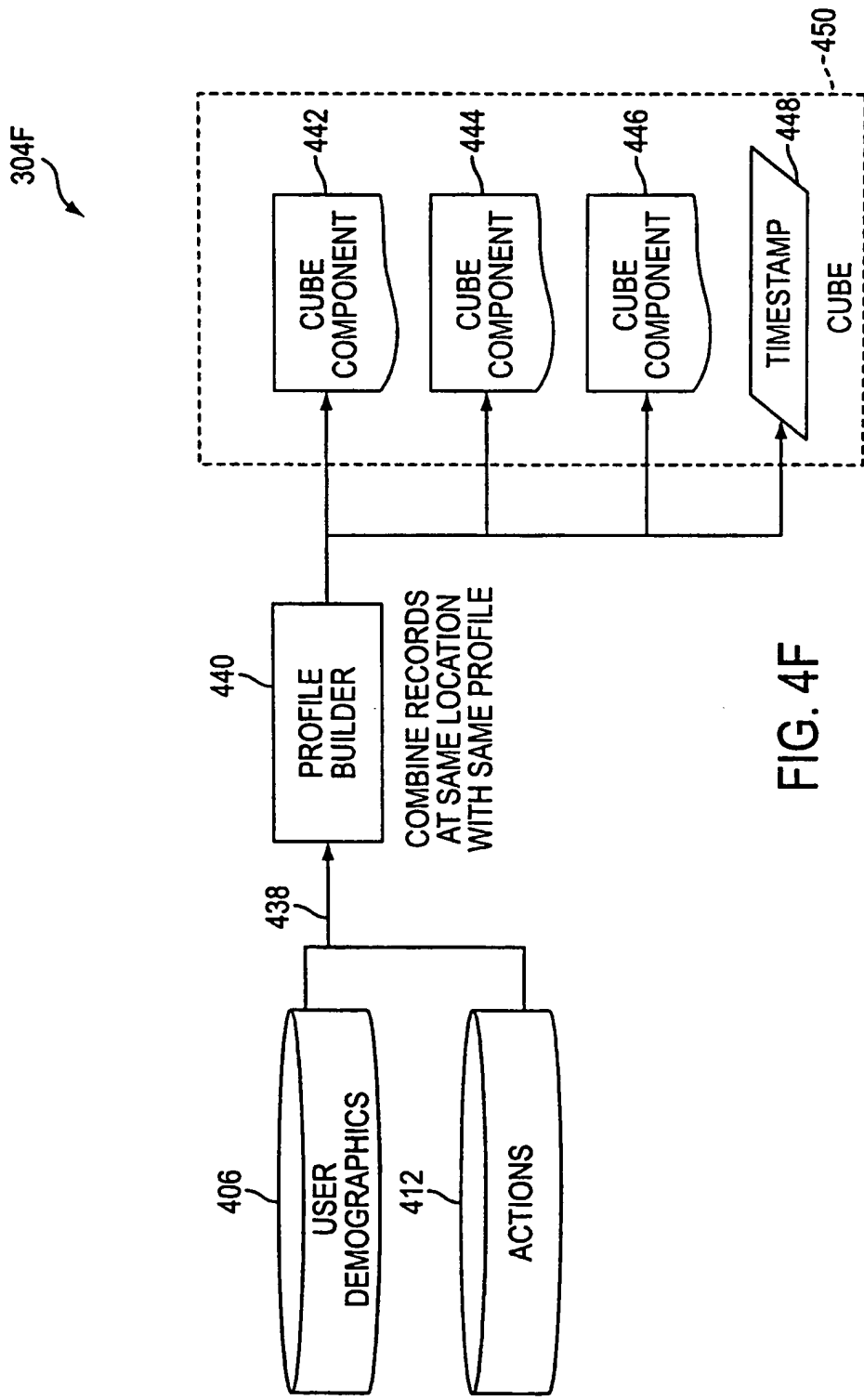


FIG. 4F

13/20

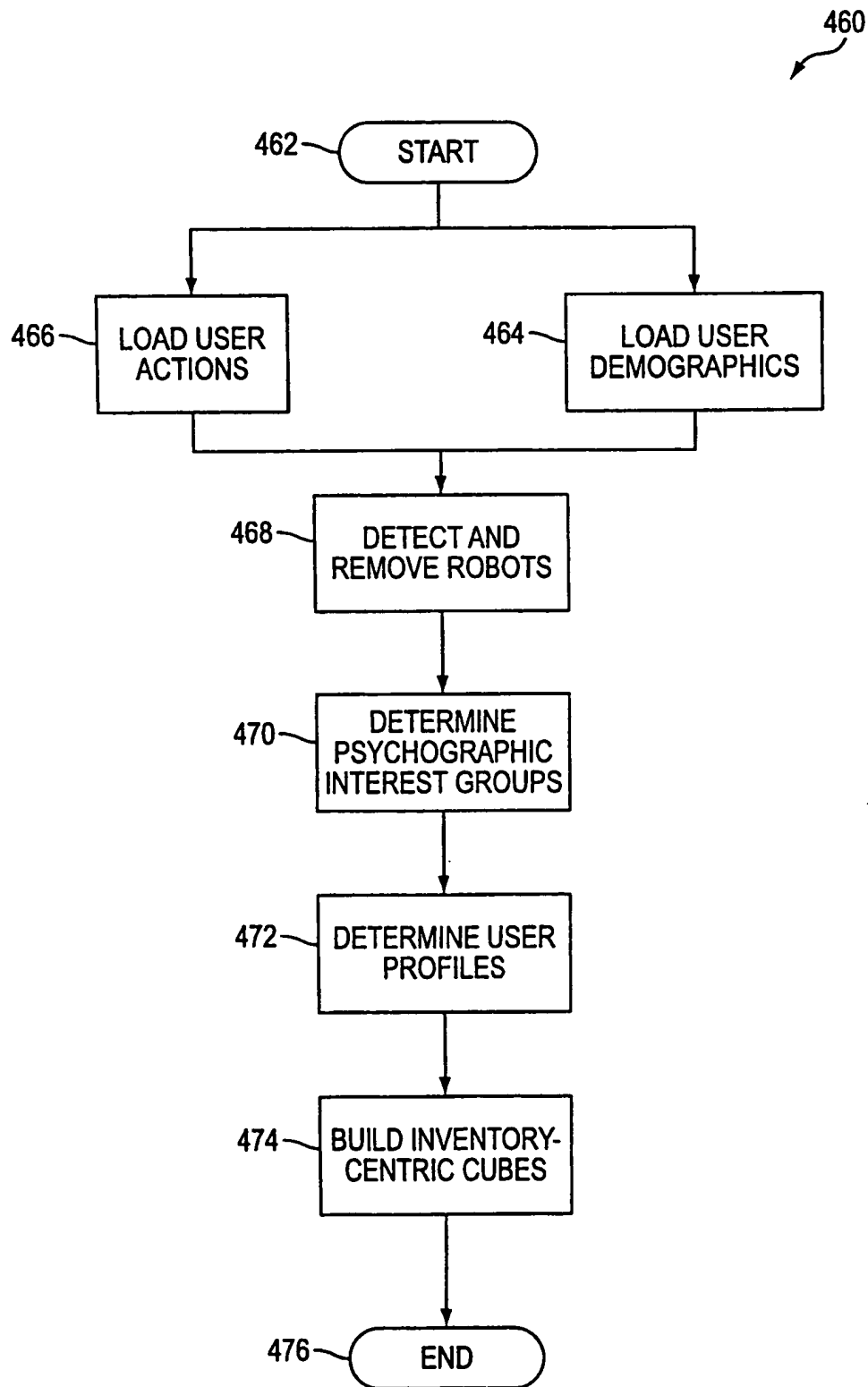


FIG. 4G

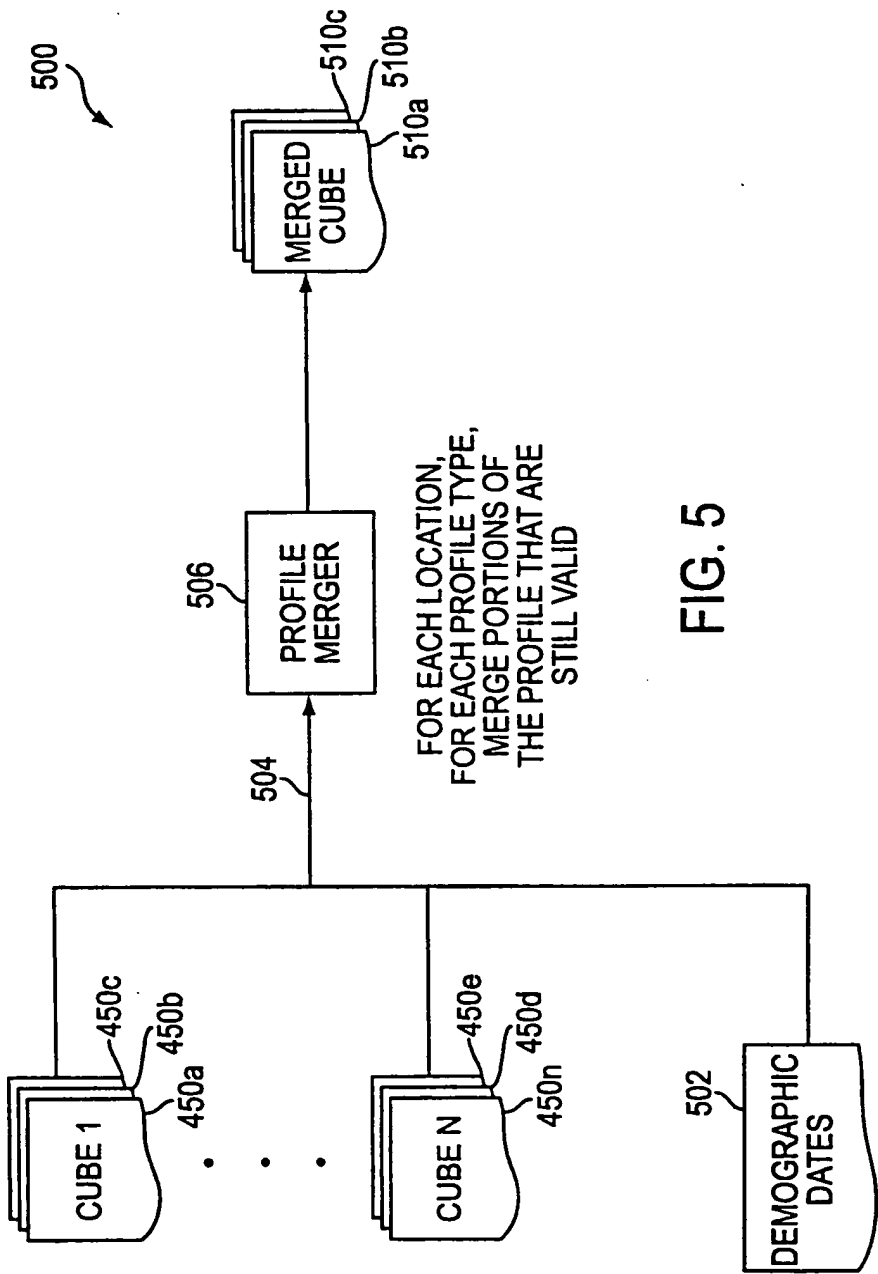


FIG. 5

15/20

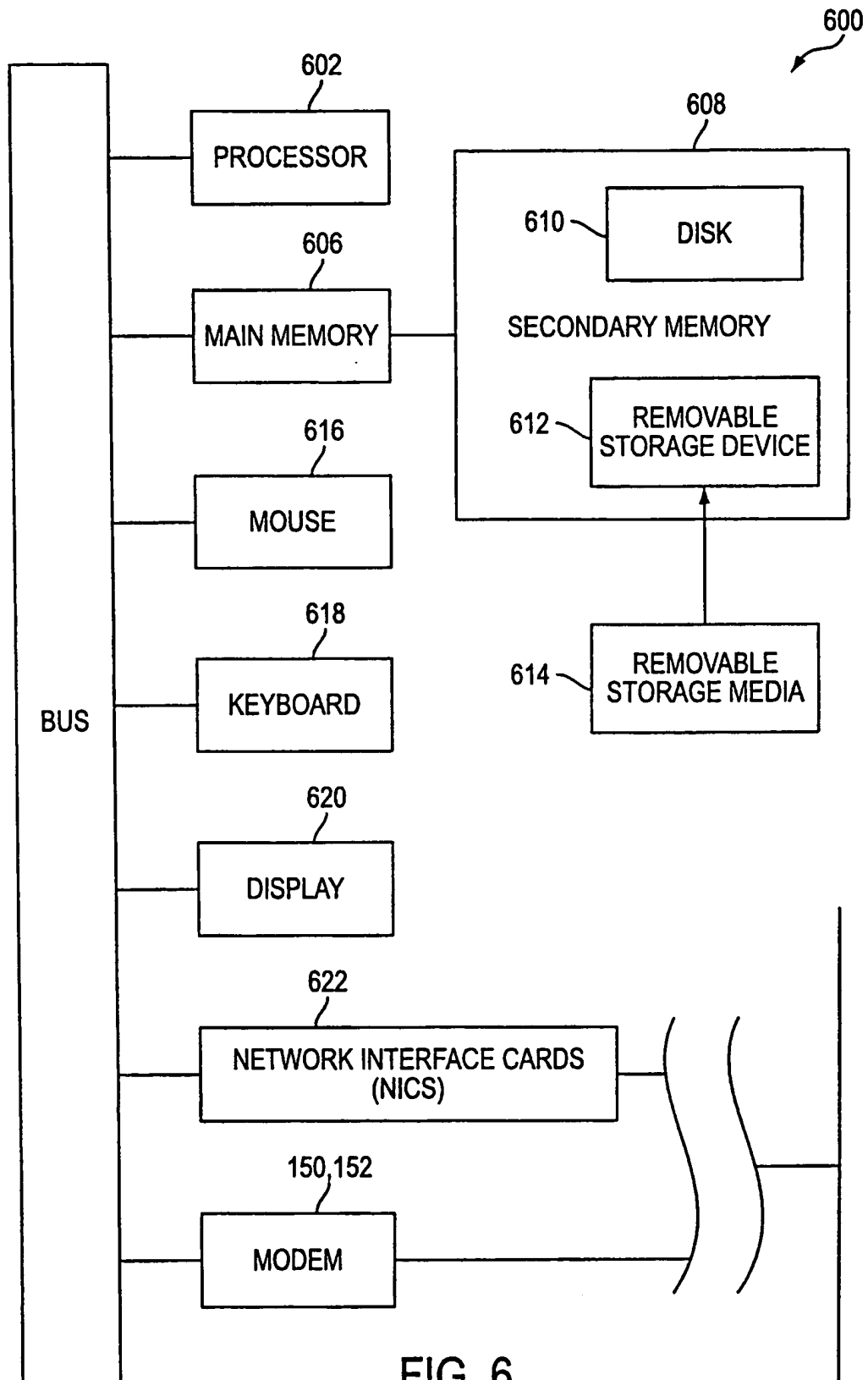


FIG. 6

808c

702

704 708

705 706 709

Traffic Report

Start 06/22/99 End 07/22/99 Weekly Apply

710 712 714

716a Location /rec/woodworking 716b

718

Include	720	Date	722	Traffic
	724	Total	726	488994
<input checked="" type="checkbox"/>	728a	6/22/99	728c	110418
<input checked="" type="checkbox"/>	730a	6/29/99	730c	110418
<input checked="" type="checkbox"/>	732a	7/6/99	732c	110418
<input checked="" type="checkbox"/>	734a	7/13/99	734c	110418
<input checked="" type="checkbox"/>	736a	7/20/99	736c	*47322
	738	Total	740	488994

742 ~ *Partial Time Window Due to End Date.

744 — Generated on 6/22/99 at 5:41 by Futuristics™ InAdvance™.

902c

904c

FIG. 7

800

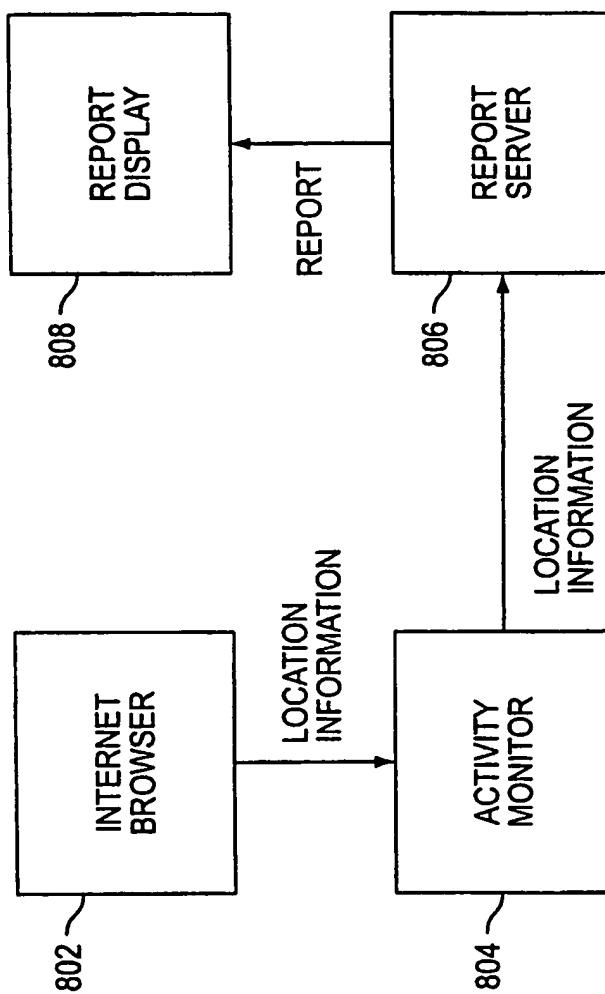


FIG. 8

808

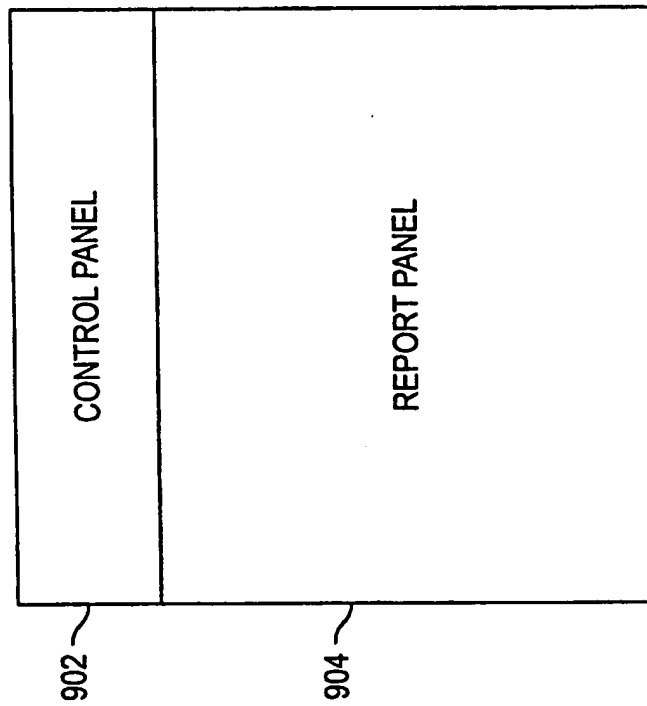
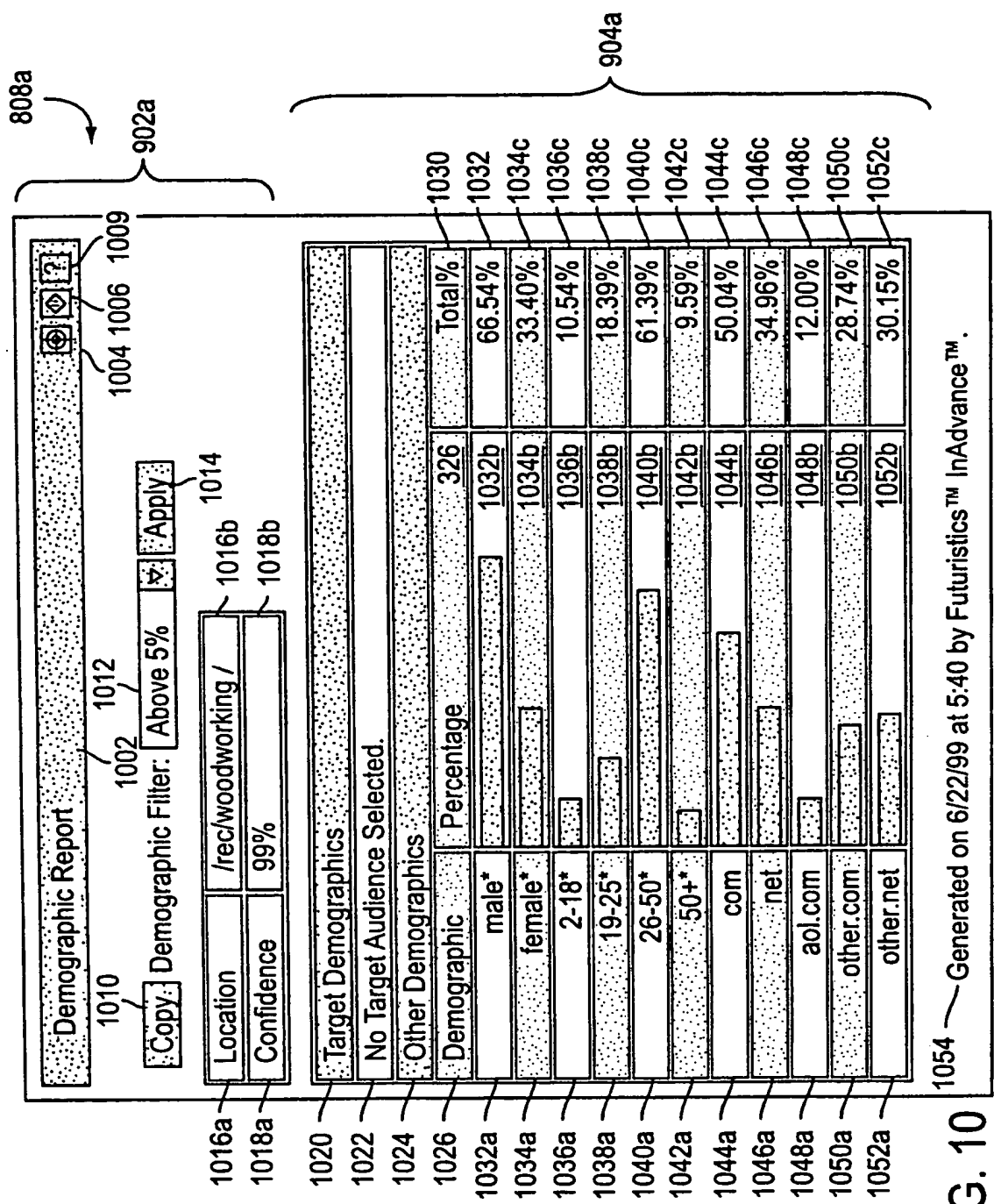


FIG. 9



1054 — Generated on 6/22/99 at 5:40 by Futuristics™ InAdvance™.

FIG. 10

20/20

